Objective Assessment of Psoriasis Treatment through Skin Images

Juan Lu

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Computing and Information Systems
THE UNIVERSITY OF MELBOURNE

May 2014

Abstract

PSORIASIS, characterised by white rough scaling and red inflamed erythema, is a common chronic inflammatory skin disease without a known cure. Since there is no consensus on the current subjective assessment methods due to the unavoidable inter- and intra- observer variances, an objective assessment of psoriasis treatment efficacy is urgently needed. The aim of this thesis is to build a severity scoring system through 2D digital skin images to evaluate the treatment efficacy of psoriasis reliably, automatically and reproducibly.

Previous research focuses on the segmentation of plaque psoriasis, the assessment of its severity and the assessment of severity changes in a short time. What is much needed is a general method for assessing the treatment efficacy of psoriasis, not only applicable to plaque psoriasis, but also applicable to other types of psoriasis for different points in time. This thesis attempts to fill the gap by considering psoriasis segmentation as erythema segmentation and scaling segmentation separately. Moreover, additional psoriasis severity features are explored and models are built for severity change analysis during a long-term treatment. This thesis contributes to:

- An erythema segmentation method is developed to segment out erythema from the psoriasis skin images. This research applies a skin model building on composition of skin pigments to do the colour analysis. The erythema segmentation accurately locates the position of erythema in digital skin images, and is not restricted to just plaque psoriasis.
- A scaling segmentation method is proposed to segment out scaling from psoriasis skin images. Scaling often comes along with erythema, but sometimes scaling appears alone. As far as we know, this is the first study to accurately detect the

position of scaling in the skin images for both situations, while previous works require the segmentation of psoriatic lesions as a first step.

- Features of erythema and scaling severity are proposed for the assessment of psoriasis severity. The erythema severity features are associated with the haemoglobin and melanin pigments; the scaling severity features are associated with the relative scaling area and the roughness degree of the scaling. The severity features correlate well with dermatologists' observation, and are shown to be better than existing methods.
- Evolution of a psoriatic lesion is evaluated in a long-term treatment. The evaluation uses a number to indicate the degree of changes of a psoriatic lesion. This research allows direct measurement of treatment efficacy in a long-term treatment rather than existing works that are only applicable to a short-term treatment.

This thesis presents the first work to reliably evaluate the treatment efficacy of a general psoriatic lesion. It shows the potential of using a computer-aided image processing system to objectively and quantitatively evaluate the psoriasis severity. It is hoped that based on this work, the future research will advance the psoriasis diagnosis as well as the treatment research.

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Juan Lu, May 2014

Publications

The work presented in this thesis has been published in the following articles:

Journal paper

- Juan Lu, Ed Kazmierczak, Jonathan H. Manton and Rodney Sinclair. Automatic segmentation of Scaling in 2-D Psoriasis skin images. IEEE Transactions on Medical Imaging. Vol. 32, No. 4, 2013, pp. 719–730.

Conference papers

- Juan Lu, Ed Kazmierczak, Jonathan H. Manton and Rodney Sinclair. A quantitative technique for assessing the change in severity over time in psoriatic lesions using computer aided image analysis. In Proceedings of the 35th Anuual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan. Jul. 3–7, 2013, pp. 2380-2383.
- Juan Lu, Ed Kazmierczak, Jonathan H. Manton and Rodney Sinclair. Automatic scoring of erythema and scaling severity in psoriasis diagnosis. In Proceedings of the 25th anniversary of the Australasian Joint Conference on Artificial Intelligence, Sydney, Australia. Dec. 4–7, 2012, pp.73–84.
- Juan Lu, Jonathan H. Manton, Ed Kazmierczak, and Rodney Sinclair. Erythema detection in digital skin images. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China. Sep. 26–29, 2010, pp. 2545–2548.

Acknowledgements

I would like to thank my supervisors, Ed Kazmierczak, Jonathan Manton and Rodney Sinclair, for their continuous support for this project. I am deeply grateful with their patient and careful guidance, which opened me the gate of academic research, and showed me the way of exploring and investigating unknown areas in the research world.

I also would like to express my appreciation to Dr. Stuart Wilson and Dr. Salem Said for their generous guidance and support in the early stage of this research. They were seldom tired of giving me their professional opinions to minor problems I met as a beginner.

I thank the support from the School of Engineering, University of Melbourne for offering me the scholarships to complete this study, and travel grants to participate international conferences.

Last but not the least, I thank my friends and my family for their support and company in this journey. I thank my friends for their boundless friendship. They are always ready to give me their help and are happy to share their experiences. In particular, I appreciate the endless care from my parents and my husband. They consistently showed their strong support even from the beginning of this study. I owe a debt of gratitude for their dedication.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1
Introduction

PSORIASIS is a chronic inflammatory skin disease. There are around 125 million people worldwide suffering from this disease, and at present there is no known cure. The most common type of psoriasis is characterised by sharply demarcated scaly and erythematous plaques on the skin (a typical example of a psoriatic lesion is shown in Figure 1.1) . Psoriasis can occur anywhere on the human body [1].

Figure 1.1: An example of a psoriatic lesion.

The prevalence of psoriasis leads to high costs of treatment for health care organisations world wide. In the United States, the direct annual cost of psoriasis treatment is estimated to be between $650 million to $2 billion. The indirect cost that takes work loss into account is as high as $11.25 billion annually. When it comes to psoriasis treatment in Australia, it is found that the cost is around $2000 individually over a biannual treatment. The chronic nature of psoriasis implies a lifetime expense.

Recently, more and more funds have been devoted into psoriasis research. A number of treatments, such as drugs, balms and radiation, have proved to be effective in control-

ling the disease. However, the treatment is very individual. Even for the same symptoms, the treatment varies among medical schools and dermatologists. In order to find out an effective solution, comparison of the treatment efficiency is necessary.

With the purpose of comparing the treatment efficacy, there are a variety of severity scoring systems that have been proposed. The scoring systems use a single number to indicate the severity degree of psoriasis. However, there is no consensus on the various scoring systems. One of the most widely accepted severity indices is Psoriasis Area and Severity Index (PASI), or the PASI score. In the PASI scoring system, the severity of a patient's psoriasis is assessed from four factors: a psoriatic lesion area and the intensity degree of erythema, scaling and thickness. However, the four factors must be estimated visually by clinicians. Despite training, intra- and inter- observer variance unavoidably occurs in practice. An objective method of assessing psoriasis severity is in demand.

In order to provide an objective assessment, a variety of devices, such as cameras, colorimeters, and 3D laser scanners, have been introduced [2–4]. These devices are used to gauge psoriatic lesions in the aspects of area, colour and thickness. Among these devices, cameras are commonly used in the psoriasis treatment to record the treatment situation, because cameras are advantageous in visual rendering, and are easy to access and use. Additionally, photography can provide a direct way of measuring the lesion area. As an example, in [2] psoriatic lesions are manually outlined in skin images to measure the percentage of affected psoriasis area in a human body. However, manual measurement is a complicated and time-consuming activity in clinical practice.

Recently, with the development of image processing technology, computer-aided image analysis has been applied in an attempt to automatically diagnose the severity of psoriasis, where the digital skin images captured by cameras are used by computer-aided image analysis systems. Most of the systems are focused on psoriasis segmentation, especially the segmentation of plaque psoriasis, though psoriasis segmentation is only an initial step for diagnosing psoriasis severity. In [5–7], psoriasis is segmented from the skin images with the help of colour information. The colour distribution of psoriasis is analysed to differentiate psoriatic lesions from normal skin. However, these techniques are not robust in uneven luminance. In [8] and [9], not only colour values but also texture

information of psoriasis are developed in the segmentation, which show an advantage in differentiating shadows and psoriasis. These methods collected individual training set from each image, but when psoriatic lesions are small, the training set is hard to identify.

By far, the only research on psoriasis severity assessment is implemented by Delgado et al. [7], where the intensity of erythema and scaling of plaque psoriasis is assessed through a special designed imaging system. In addition, Delgado et al. [10] assess the change of psoriasis through a image subtraction technology.

Even though plaque psoriasis is a common type of psoriasis disease, it is hard to apply these methods to other types of psoriasis, because of the diversity in psoriasis appearance. Moreover, the developed severity change analysis technology is only available for a short-term treatment diagnosis. When the treatment spans more than a month, changes of lesion content and boundary make the method unsuitable.

The ultimate goal of this research is to build a severity scoring system through 2D digital skin images to evaluate the treatment efficacy of psoriasis reliably, automatically and reproducibly, and that can be used to compare lesions over long time spans.

1.1 Contributions

Previous research focuses on the segmentation of plaque psoriasis and the assessment of severity in plaque psoriasis. What is much needed is a general method for assessing the treatment efficacy of psoriasis, not only applicable to plaque psoriasis, but also applicable to other types of psoriasis. This thesis attempts to fill the gap by considering psoriasis segmentation as erythema segmentation and scaling segmentation separately, due to the fact that erythema and scaling are common symptoms and compose a general psoriatic lesion. Moreover, psoriasis severity features are explored to improve the existing severity assessment methods, and models are build for the severity change analysis during a long-term treatment. By doing this, the treatment efficacy can be assessed for a wide range of psoriasis based on different lesions or different time points. Our findings can be divided into two stages: the psoriasis segmentation stage and the psoriasis assessment stage. Specifically, this thesis has the following contributions.

1. In the psoriasis segmentation stage, the erythema segmentation is developed to segment out erythema from the psoriasis skin images. This research applies a skin model building on decomposition of skin colour into its melanin and haemoglobin components. The erythema segmentation method accurately locates the position of erythema in digital skin images, and is shown to outperform the previous method [11], especially with respect to mild erythema.
2. Development of a scaling segmentation method to segment out scaling from psoriasis skin images. Scaling often comes along with erythema, but sometimes no erythema appears in the area surrounding the scaling. As far as we know, this is the first study to accurately detect the position of scaling in the skin images for both situations. This scaling segmentation does not require the segmentation of psoriatic lesions as a first step as conducted in the previous research [7].
3. In the psoriasis severity assessment stage, erythema severity features and scaling severity features are proposed. The erythema severity features are associated with the haemoglobin and melanin quantities in the skin; the scaling severity features are associated with the relative area and roughness degree of the scaling. The severity features correlate well with dermatologists' observations, and are shown to be better than existing methods. There is no accepted level of accuracy for psoriasis assessment. The accuracy of our severity assessment is above 75%, which satisfies the general requirements of accuracy in biomedical engineering area [12,13].
4. Evolution of psoriatic lesions is evaluated in long-term treatments. The evaluation uses a number to indicate the degree of change of a psoriatic lesion. Since psoriasis is a chronic skin disease, this method allows direct measurement of treatment efficacy in a long-term treatment rather than existing works that are only applicable to a short-term treatment [10,14].

1.2 Thesis organisation

Psoriasis and its clinical severity assessment are introduced in Chapter 2, followed by a review of devices investigated for the objective severity assessment. This review shows

a competitive advantage of cameras over other investigated devices, e.g. ultrasound, microscopes, spectrometers et al. Next, this chapter presents a survey of existing computer-aided diagnosis systems based on image analysis, and the applications of computer-aided diagnosis systems for psoriasis treatment evaluation.

In Chapter 3, a method for erythema segmentation is presented. This chapter begins with a description of skin region identification in psoriasis skin images. It is followed by an analysis of a skin colour decomposition, through which a skin colour is decomposed into its melanin and haemoglobin components. In the following section, an application of a Support Vector Machine (SVM) to identify the locations of erythema pixels is presented. Finally, parameters used in the skin colour decomposition are validated, and the superiority of the performance of the proposed method is shown by comparing with a nearest neighbour classification method in the experiment.

Chapter 4 presents an algorithm for scaling segmentation. The scaling segmentation algorithm does not rely on using pre-identified psoriasis to identify regions of scaling, as conducted in previous researches [7]. Two scaling features are introduced at first: one is the scaling contrast map that enhances the contrast between scaling and its surrounding erythema, and the other is the Gabor texture that differentiates rough scaling from smooth normal skin. This is followed by a presentation of the proposed semi-automatic training set collection and the proposed SVM-based Markov Random Field (MRF) classification to identify location of scaling pixels. In the experiment, the performances of the training set collection and the SVM-based MRF classification methods are analysed, and they are respectively compared with existing clustering methods, k-means and fuzzy c-means, and classification methods: SVM and MRF. The results show better performance of the proposed methods. This is followed by a parameter estimation to validate parameters used in the algorithm.

Building on the results from Chapters 3 and 4, Chapter 5 presents the algorithms of erythema severity scoring and scaling severity scoring. This chapter begins with a presentation of segmenting the whole psoriatic lesion through the erythema segmentation and the scaling segmentation. The severity features and the severity classifiers are introduced in the following section. Finally, in the experimental section, it shows that the

severity features are strongly correlated with scores given by dermatologists, and the severity scores given by the severity classifiers are reliable.

The severity change analysis is presented in Chapter 6. A multiple linear regression model is introduced to analyse the relationship between the change of severity features and the change of severity scores with respect to the erythema severity changes and the scaling severity changes. This is followed by an experimental section showing that the change of severity features and the multiple linear regression models are well correlated with dermatologists' diagnoses.

Chapter 7 is a summary of this thesis. Contributions of this work are reviewed followed by a discussion of the limitations as well as future research with respect to improvement of this work and potential applications to other skin disorders.

Part II

Background

Chapter 2
Literature Survey

This chapter gives an introduction to psoriasis, including psoriasis severity scoring, followed by a review of devices previously investigated for objective assessment of psoriasis severity. Next, this chapter provides an overview of existing computer-aided diagnosis systems based on digital image processing, including relevant statistical machine learning techniques used in these systems. This is followed by a review of existing applications of computer-aided diagnosis systems for psoriasis severity scoring. This chapter is concluded with a description of the problems resolved in the thesis.

2.1 Psoriasis

Psoriasis is a chronic skin disease, which manifests as red, itchy and scaly patches of skin. It is genetic and results from abnormal activity in the body's immune system. Other factors, such as diet, stress and alcohol consumption may trigger the onset of psoriasis. However, its pathogenesis is still not fully understood, and many researchers are working on the problem [15,16]. UV radiation and topical agents, including salicylic acid, steroids, and Vitamin-D analogues, are often applied to treat psoriasis [16]. Though remission can be achieved by the treatment, psoriasis is not currently curable.

Psoriasis primarily affects knees, elbows, scalps, hands feet and lower backs. Men, women, children and even newborn babies can have this disorder, which unfortunately can be life-long [15]. Additionally, psoriasis causes pain and itching, and in some rare cases, can be life-threatening. Moreover, since psoriasis affects the patients' physical appearance, it often dramatically negatively impacts the patients' social lives.

2.1.1 Types of psoriasis

The symptoms of psoriasis include erythema (inflamed red skin), scaling (thickened patches with silver scales) and swelling. It may affect a few regions of the skin, may be moderate, or may widespread. Based on morphology, distribution and pattern, there are five types of psoriasis.

1. Plaque psoriasis is raised red-looking skin lesions covered by silvery scaly skin. It is the most common type of psoriasis, affecting 80-90% people.
2. Pustular psoriasis is the second most common type of psoriasis. It begins with inflamed reddened skin that is followed by scaling and raised bumps or blisters.
3. Guttate psoriasis is characterised by numerous small red spots with very fine scaling. It appears over a large area of a body
4. Flexural psoriasis is dominated by smoothed inflamed patches of skin. It often occurs in skin folds, such as arm-pits and under breasts.
5. Erythrodermic psoriasis appears as widespread inflammation and exfoliation of the skin. Sometimes, swelling appears at the same times. The scaling is more fine and flaky than for plaque psoriasis.

In Figure 2.1 samples of these types of psoriasis are shown. Often, psoriatic lesions of different types occur alone. Sometimes, there is an overlap or transition from one type to another due to various triggers or the evolution of the disease [15].

2.1.2 Economic impact of psoriasis

Psoriasis is a skin disease. Reports indicate that about 3% to 4% of the population worldwide suffering from this disease, and in Australia the number is around 2.6% [17].

The economic impact of psoriasis is correspondingly high. The overall annual cost of caring for individuals with clinically significant psoriasis has been estimated to be from $650 million to $2 billion in the United States [18]. Social morbidity brought by psoriasis causes a considerable economic impact as well. Patients with psoriasis experience emotional suffering related to their appearance, limitation of activity, and loss of time for employment because of the need to receive treatment [1, 19]. Total direct and indirect

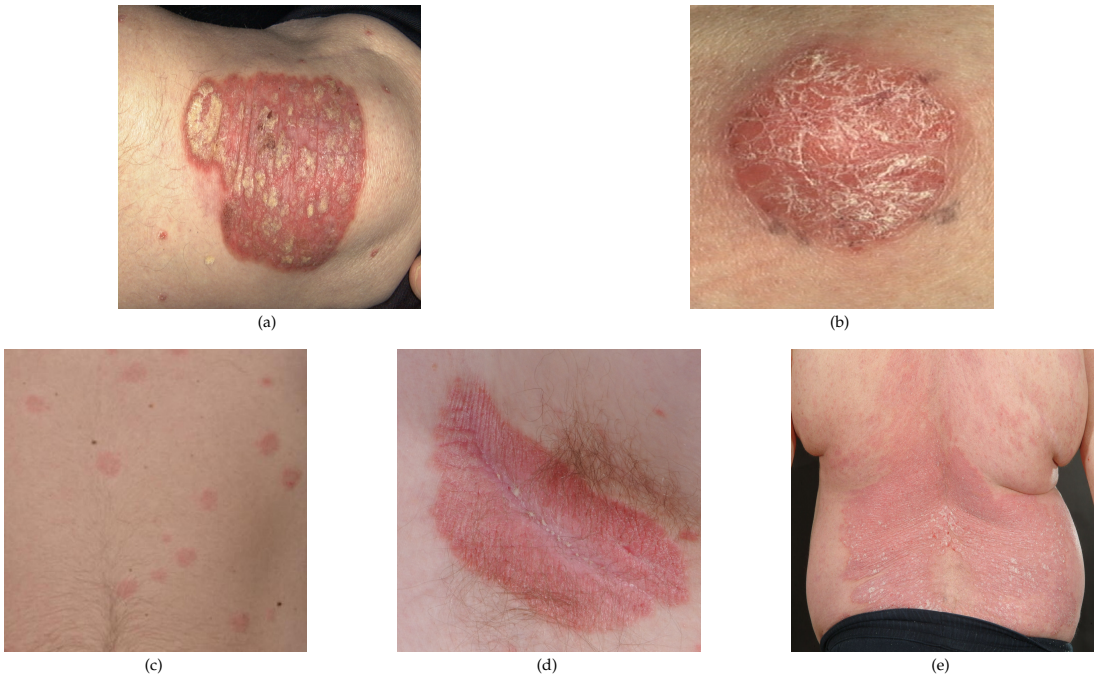


Figure 2.1: Sample images of different types of psoriasis. (a) Plaque psoriasis; (b) Pustular psoriasis; (c) Guttate psoriasis; (d) Flexural psoriasis; (e) Erythrodermic psoriasis.

health care costs of psoriasis for patients are calculated at $11.25 billion annually, with work loss accounting for 40% of the cost burden [1].

In Australia, according to the research conducted by Jenner et al., individual expenditure in this disease can be as high as AUD $2000 over a 2-year treatment period [19]. The chronic nature of the disease implies that patients usually spend a lifelong expense for this care.

2.1.3 Psoriasis severity assessment

Due to the lack of knowledge about the causes of psoriasis, treatment must be individualised. Dermatologists typically use different treatments to treat psoriasis with the same symptoms. Therefore, to assess which treatment is comparatively effective, we need to assess treatment efficacy. The efficacy of psoriasis treatment is evaluated through a severity score. A severity scoring system translates a clinical judgement into a number in a numeric scale to quantitatively and qualitatively measure psoriasis severity. A severity

scoring assists dermatologists to predict the treatment outcome and intervene to achieve an expected outcome.

However, there is no consensus on the psoriasis severity scoring systems. According to the guidelines issued by the European Medicines Agency, ideally a psoriasis severity assessment should involve the body surface area affected by psoriasis, the intensity of local symptoms (erythema, scaling and elevation), history of previous treatments, disease duration, degree of disability and the impact of the disease on patient's quality of life [16]. Several popular severity assessment methods used in practice are introduced in the subsections below.

Psoriasis area and severity index

Psoriasis Area and Severity Index (PASI) is one of the most widespread methods [20] in clinical treatment and research. In [21] Naldi reviewed 44 scoring systems used in 171 clinical trials of psoriasis therapies and observed that the PASI scoring was used in about half the trials. PASI scoring was proposed by Fredricksson and Pettersson in 1978 for use in a single clinical trial, and subsequently became popular. PASI gives a single index that captures the severity in four different body regions: head (occupying 10% of total body surface), upper limbs (20%), trunk (40%), and lower limbs (30%), each of which is weighted based on the proportion to the whole body surface area. In every region, the affected area is graded on a 0-6 scale corresponding to the percentage of coverage. Additionally, the severest psoriatic lesion is picked out in each body part and is rated on a scale of 0-4 by assessing the three symptoms visually: redness, thickness and scaliness. The final score of psoriasis severity is the sum of the scores for the three symptoms, together with the area affected based on the weighting for each region. The equation of the PASI scoring is expressed below:

$$\begin{aligned} \text{PASI} =& 0.1(R_h + T_h + S_h)A_h + 0.2(R_u + T_u + S_u)A_u \\ &+ 0.3(R_t + T_t + S_t)A_t + 0.4(R_l + T_l + S_l)A_l \end{aligned} \tag{2.1}$$

where R_i, T_i, S_i and A_i are the redness, thickness, scaliness and area scores for different regions of the body, where the subscript $i \in \{h, u, t, l\}$ indicates the regions head (h), upper limbs (u), trunk (t) and lower limbs (l) respectively. The range of PASI score is from 0 to 72, and this score is discrete, with increments of 0.1 value [22].

Psoriasis global assessment

Another frequently used assessment method is Psoriasis Global Assessment (PGA). This method divides the severity into seven scales and gives the severity a global description for each scale. The PGA method is more easily used than PASI in clinical practice, is more intuitive and does not require quantitative calculation. However, this method is still a visual scoring one. The severity score is subjectively assessed by dermatologists. In Table 2.1, an example of the PGA calculation is shown.

Severity	Description
Clear	No signs of psoriasis
Almost clear	Intermediate between mild and clear
Mild	Slight plaque elevation, scaling and/or erythema
Mild to moderate	Intermediate between moderate and mild
Moderate	Moderate plaque elevation, scaling, and/or erythema
Moderate to severe	Marked plaque elevation, scaling, and/or erythema
Severe	Very marked plaque elevation, scaling and/or erythema

Table 2.1: An example of the PGA calculation.

The lattice system

Many other severity assessment methods have been proposed recently. Among these, the lattice system is another popular method. This scoring process of the lattice system is similar to PASI. It quantitatively takes into account the percentage of affected area and the severity of symptoms assessed on a scale of 0-4. The severity result is divided into eight categories from clear to very severe. It attempts to provide a better inter- and intra-observer variances than PASI [23].

Quality of life questionnaires

Besides the aforementioned clinical severity assessment, Quality of Life (QOL) questionnaires are increasingly used in clinical research and routine clinical practice, due to emphasis on evidence-based severity scoring. In the questionnaires, patients' feeling and the influence that psoriasis has on their lives are investigated. A psoriasis severity score is calculated based on answers of the questionnaire. In [20], eight QOL questionnaires are investigated and compared, including the widely used Dermatology Life Quality Index and the Medical Outcome Survey Short Form 36. The QOL questionnaires are usually used as a complement to the clinical severity assessment. However, reliability and reproducibility are still problematic in these questionnaires. None of the QOL questionnaires are recommended for psoriasis assessment.

Among the variety of severity scoring systems, the PASI scoring is still the dominant one. The severity factors considered in the PASI, that is, the affected area, and the severity of lesions assessed in the respects of redness, scaling and elevation, are all included in other clinical severity scoring systems as well. The PASI provides a prototype for the severity scoring [23], while other clinical severity scoring systems work on the prototype with various adaptions in an attempt to avoid the biases in PASI. The severity factors are either rescaled or recombined in the severity scoring systems. However, the advantages of the other scoring systems are still controversial.

The following weaknesses prevent any of the scoring systems from being universally accepted:

1. The scoring results are not objective. The symptom and area severity scores are visually determined by dermatologists. The severity index depends on dermatologists' experience. Therefore, the inter-observer and intra-observer variances are unavoidable.
2. The methods are not accurate. In the PASI scoring and the lattice scoring system, the severity of psoriatic symptoms and the area of psoriatic lesions are coarsely rated. When it comes to the PGA scoring system, the definitions of "mild", "moderate", and "severe" severities are not clear. Moreover, the weights assigned to the body parts in Eq. (2.1) of PASI scoring are improper to accurately reflect the

clinical severity.

3. The scale numbers and the scoring rules of the visual assessment methods are not sensitive to small severity changes. Additionally, in the PASI scoring, the small percentage of body covered by psoriasis limits the sensitivity to the severity changes of the lesions.
4. Calculation of the popular PASI is complicated. It is too time-consuming to use in busy clinical practice.

Thus, there is a need to objective assess the severity of psoriasis.

2.2 Devices Non-automatically Applied for Objective Assessment of Psoriasis Severity

Due to the limited objectivity of the aforementioned subjective scoring methods, a number of devices are investigated to evaluate psoriasis severity more objectively. This section focuses on a review of devices that are used non-automatically for assessment of the severity of psoriasis. Using these devices, psoriasis is measured from the aspects of area, colour and thickness.The investigated devices are described below.

2.2.1 Area measurement

The most widely used devices for psoriatic area measurement are cameras. As far as we know, the earliest research of using cameras to assess the affected area of psoriasis is found in [24], where patients with psoriasis are photographed with a film camera and the affected area of psoriasis is measured by computer-assisted planimetry. As the development of optical imaging technology, digital cameras are used to photograph patients with psoriasis, and area of psoriasis is measured by computer aided image analysis [2, 25, 26]. In [2], dermatologists outlined the psoriatic lesions and body parts on separate transparent sheets overlaid on photographed images of psoriasis patients. After the transparent sheets are digitised and the outlined regions are filled on a computer, area of the marked lesions and body parts is quantitatively described by pixel counts.

In [2, 24–26], the results of area scoring in the PASI score system are compared with

results derived through the use of cameras, and advantages of using cameras for area measurement are proved. In [2, 24, 25], it is pointed out that even experienced clinicians may differ in the estimates of the body area. In the experiment conducted in [2], it shows that a visual estimation of PASI scoring tends to be overestimated compared with the usage of cameras. Additionally, [26] and [25] prove that using cameras is able to measure the area changes that are not detected by the subjective estimation.

2.2.2 Colour measurement

Erythema colour is related to the degree of inflammation and reflects the severity of erythema in psoriasis. In order to objectively measure erythema severity, spectrometers and colorimeters are applied.

A spectrometer measures the wavelength of light reflected by skin. As far as we know, the earliest spectrometer designed for the measure of erythema severity was proposed in [27]. The erythema severity is measured based on the absorbance of two primary skin pigments: melanin and haemoglobin. The absorbance of melanin is called the melanin index, and the absorbance of haemoglobin is called the erythema index [27, 28]. Though both erythema index and melanin index can be used to measure the severity of erythema, erythema index is more sensitive to the erythema severity [28].

A colorimeter measures reflecting light from skin through a tristimulus colour analysis [29]. A colour is measured by giving relative quantities of three primary colours that are defined by the CIE standard [30]. The most commonly used colour space that is derived from the CIE colour space is the $L^*a^*b^*$ colour space. To our knowledge, colorimeters are first proposed to measure the severity of erythema in 1990 [31]. In [3] and [25], colorimeters are used to record erythema colour by averaging five-point measurements. These five points are randomly taken within the lesion. In [32], hue values of skin are derived from colour values in the $L^*a^*b^*$ colour space measured by a colorimeter, and the difference of hue values between erythema and normal skin is used to analyse erythema severity.

Spectrometers and colorimeters measure the colour by providing their own illumination sources, so that the colour value is not affected by the change of illumination from

the surrounding environment. However, since the erythema colour is randomly sampled for the measurement, when the area of psoriasis is large, these methods are not reliable.

2.2.3 Thickness measurement

The reason that psoriatic lesions are often thicker than normal skin is proliferation of epidermis and swelling. There are three ways investigated to measure the psoriasis thickness:

1. Imaging skin with ultrasound, and Optical Coherence Tomography (OCT).
   Ultrasound technology and OCT are techniques that can be used to render subcutaneous layers non-invasively. By using these techniques, epidermal thickness is measured from the skin surface to the derma-epidermal junction. Over the past two decades, ultrasound technology is increasingly applied [3, 25, 33]. In [3, 25], ultrasound was used to measure the depth of psoriasis by randomly sampling within the lesion and reporting the thickness as the average value of the random samples. Recently, OCT is applied in [34] to measure the epidermal thickness by imaging inner skin structure in real time.
2. Examining histopathological images.
   Histopathological images of psoriasis are obtained through a biopsy method, and are observed under microscope. The epidermal thickness is determined by measuring the distance between the top and the bottom of rete (a skin structure that extends from epidermis to dermis). In [35], five samples are randomly sampled from a histopathological image of psoriasis, and the epidermal thickness is the average value measured using the five samples.
3. Imaging skin with 3D laser scanners.
   3D laser scanners are also used for the measurement of psoriasis thickness [4]. Through the use of this device, only depth information of the skin surface is measured. Psoriasis thickness is measured from the lesion surface to the lesion base. This method converts the thickness information accessed by touch to an accurate number. Disturbances from the inner skin, such as fat layers, are avoided. Meanwhile, this method is able to assess the thickness of psoriatic swelling directly.

In Table 2.2, a comparison of the devices surveyed for the objective assessment of psoriasis severity is presented. Though the severity of psoriasis can be measured with these devices objectively and quantitatively, the use of these device is not automatic. Their use can be time consuming and so prevents their wider usage.

Among the devices in Table 2.2, cameras are commonly used in the psoriasis treatment to record the treatment situation, because cameras have the advantage of rendering lesions visually. This advantage is proved in [36], where a good correlation is found between the PASI scores assessed from psoriasis patient images and the scores diagnosed in patients' visits. Moreover, the easy access and use of cameras further boost the application of cameras in the psoriasis treatment.

When computer-aided diagnosis systems are introduced into psoriasis severity assessment, using psoriasis images taken by cameras, the automatic assessment of the efficacy of psoriasis treatment, becomes possible. With the development of image processing and machine learning techniques, not only psoriasis area, but also the severity of erythema and scaling can be assessed quantitatively.

Table 2.2: Devices for assessment of psoriasis severity.

Devices	Measurement	Principle	Advantage	Disadvantage
cameras [2, 24–26]	area	count pixels in outlined psoriatic lesions	correct area bias from subjective measurement	manually outlining lesions is required
colorimeters, spectrometers [3, 25, 27, 28]	colour	measure light reflectance to estimate absorbance of erythema	robust to illuminance changes of outside environment	only colour measurement of local area is available
microscopes [35]	thickness	measure distance between top and bottom of ret in histopathological image	directly assess the thickness	an invasive method: psoriatic lesions need to be damaged
ultrasound, OCT [3, 25, 34]	thickness	suffered accuracy from inner skin structure, such as fat layer	non-invasive method	measure distance from skin surface to dermal-epidermal junction
3D Laser scanners [4]	thickness	measure distance from lesion surface to lesion base	measure thickness directly and non-invasively	suffered poor calibration accuracy for large area

2.3 Computer-aided Diagnosis Systems Based on Digital Image Processing

Before a survey of existing computer-aided diagnosis systems for efficacy assessment of psoriasis treatment, we give a review of generally defined computer-aided diagnosis

systems based on digital image processing. The concept of computer-aided diagnosis systems was proposed in the 1980s [37]. A computer-aided diagnosis system analyses images captured with a variety of imaging modalities to assistant clinical diagnosis. It is pointed out that a computer-aided diagnosis system makes a complementary diagnosis to help physicians to get the correct one [37]. When the decision is hard to make, the diagnosis provided by the computer would serve as a strong reference.

2.3.1 The general elements of a digital image diagnosis system

We give a model of pixel based image processing diagnostic systems for the purpose of the discussion in Sections 2.3.2-2.3.4. The frame of this model is presented in Figure 2.2.

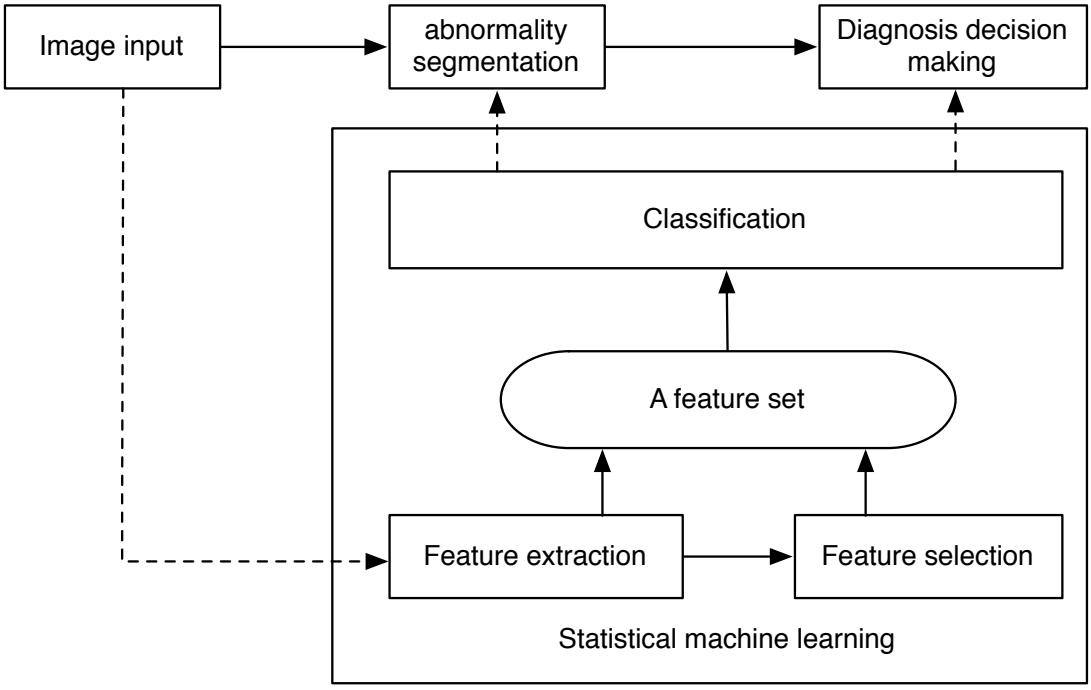


Figure 2.2: A general frame of pixel based image processing diagnostic system.

There are many existing examples of this model in the literature [38–49]. In [49], prostate cancer is diagnosed by segmenting prostate pixels from non-prostate pixels in magnetic resonance images. In [40], a computer-aided diagnosis system is designed to quantitatively assess the expression of the HER2 gene using digital microscopy. This

system includes a colour pixel classifier used for the segmentation of epithelial cell nuclei and membrane, and a membrane staining assessment part based on the segmentation result.

The model of the computer-aided diagnosis system is composed of four components: (1) image input, (2) abnormality segmentation, (3) diagnosis decision making, and (4) the usage of statistical machine learning techniques for the abnormality segmentation and diagnosis decision making. In the image input part, digital images captured by various imaging modalities are inputted into computers for analysis. The abnormality segmentation component identifies the localisation of an abnormality in the images. The localisation knowledge is then used in the following diagnosis decision making part to give an evaluation of the abnormality.

Since both abnormality segmentation and diagnosis decision making can be modelled as classification problems, statistical machine learning techniques can be used for the classification. Specifically, in the disease segmentation part, the classification problem is to differentiate "abnormal" pixels from "normal" pixels, while in the diagnosis decision making part, the classification of "benign" and "malign" diseases, or an assessment of the severity together with a description of the diagnosis is required. Statistical machine learning techniques play an important role in computer-aided diagnosis systems.

As illustrated in Figure 2.2, there are three modules in a statistical machine learning system: feature extraction, feature selection and classification. A feature set is a set of feature components that characterise the current classification problem. The feature set can be derived from the feature extraction and feature selection modules. The feature set provided by the feature extraction module consists of the raw features directly extracted from images. By using the raw features, the feature selection module selects more relevant features. In the classification module, boundaries of different classes are localised by optimising a decision function in a feature space constructed by the feature components. The remaining of this section reviews existing applications of the feature selection, feature extraction and the classification techniques in computer-aided diagnosis systems.

2.3.2 Feature extraction

Feature extraction generally derives features according to the classification guidelines of the current problem. The derived features are called the raw features, and are quite different between the abnormality segmentation and the diagnosis decision making problems.

For the abnormality segmentation task, directly extracted features from images include colour and texture. Colour features are descriptors of the colour information digitised by an imaging modality, and texture features describe regional colour intensity changes.

A colour feature is often extracted from a colour space, e.g. the RGB space, the $L^*a^*b^*$ space, and the HSV space. The RGB space, describing colours in terms of redness, greenness, and blueness, is a colour space that an imaging device directly works on. The $L^*a^*b^*$ space uses a luminance component and two colour components, redness to greenness and blueness to yellowness, to describe colours, and is perceptually linear with human visual explanation. The HSV space is created by mimicking painting. Here, H is for hue, S is for saturation and V is for the value of chroma. The above mentioned colour spaces may work separately or together in an image segmentation algorithm. In [50], the $L^*a^*b^*$ colour space is explored for the segmentation of gland units in microscopy images of the prostate tissue. [51] uses the G value from the RGB space, the H value from the HSV space and the a value from the $L^*a^*b^*$ space for the segmentation of hemangioma from skin images. For the diagnosis of breast cancer, [40] works on miscopy images of breast tissue and segments epithelial cell nuclei and membrane from background by using R, G, and B values in the RGB space, S and V values in the HSV space, and a and b values in the $L^*a^*b^*$ space.

By characterising the surface texture of an abnormality, a texture feature plays another important role in abnormality segmentation. There are two kinds of methods to analyse the texture: statistical based methods, and filtering based methods. The statistical based methods, such as the co-occurrence matrices, statistically measure the texture pattern [52]. The filtering based methods, including Fourier transformation [53, 54], wavelets filters [55] and Gabor filters [56], analyse filter responses of an image through the use of a convolution operation. Among the texture analysis methods, Gabor filters

receive a constant attention because of their high correlation with human perception. Gabor filters are used in our scaling segmentation technique (see Chapter 4 for the detailed description).

When feature extraction techniques are used in the diagnosis stage, features are descriptors related with the diagnosis guidelines. For example, in [41] three sets of features, the perimeter, the ratio of the major to minor axis, and the mean of the grey level intensity, are extracted in histopathological images to diagnose whether breast cancer is benign or malign. The three features are respectively corresponding to cell size, shape and nucleoli appearance, which construct the basis of the clinical breast cancer diagnosis.

2.3.3 Feature selection

The feature selection component attempts to identify more relevant features by analysing the features derived by feature extraction. Feature selection aims to reduce the redundant information in the extracted raw features. This method is helpful for the computational complexity and the classification accuracy. Popular feature selection methods include Principle Component Analysis (PCA) and Independent Component Analysis (ICA).

PCA transforms the raw features into uncorrelated features. Transformed features with large variance are selected, and hence the dimension of the feature space is reduced. PCA has been used for the feature selection in both medical image segmentation and image analysis-based diagnosis. For the example of medical image segmentation, in [57], PCA is used to segment the optic disc from a fundus image. The component with the largest variance is extracted in the RGB colour space to construct a grayscale image for the segmentation analysis. In [58], a grayscale tooth image is filtered with a bank of Gabor filters. The resulting images are stacked and analysed with PCA. The most signification component is used in the following segmentation analysis. For the example of image analysis based diagnosis, [42] extracts three kinds of features from the segmented white blood cells. The most relevant feature is selected using PCA and is used in a neural network classifier to recognise the type of a white blood cell.

ICA transforms the feature components into independent ones. This condition is stronger than PCA. The independent components are not only mutually uncorrelated,

but also statistically independent, while the uncorrelated principle components are not equivalent to the independent components. The ICA analysis explores deeper hidden information than PCA. Application of the ICA analysis exists in both the segmentation stage and the diagnosis stage. In [59], the ICA analysis works on a feature set derived by a bank of filters to improve exclusivity of the feature set in iris segmentation. In [60], ICA is used to blindly extract essential features to guide the diagnosis of breast cancer from multiple spectral images. In [61], ICA is applied to the MRI images to derive the independent basis functions for the diagnosis of Alzheimer's disease.

Details of the PCA and ICA analysis are described in the feature selection part of the erythema segmentation (see Chapter 3).

2.3.4 The classification problem

There are two kinds of classifiers widely used in the biomedical engineering. One is the sample based classifiers that perform the classification based on features of individual samples. This kind of classifiers can be used in the abnormality segmentation and the diagnosis decision making. Typical used sample based classifiers include K-Nearest Neighbours (KNN), decision trees, Linear Discriminant Analysis (LDA), Multilayer Perceptrons (MLPs), and Support Vector Machines (SVMs). The other kind is the graph based classifier. This kind of classifiers models image coordinates as vertices in graphs and takes the image structure into consideration. The graph based classifiers are vastly used in the segmentation of abnormalities.

Sample based classifiers

KNN is a classification method that identifies the class of a test sample as the class that most k nearest training samples belong to. The Euclidean distance is usually used to determine the distance of training samples to the test sample. The class of the test sample is the class that most k nearest training samples belong to. KNN is mainly used in the diagnosis decision making stage, such as pulmonary embolism detection [62], breast cancer detection [63], and coronary calcification scoring [64, 65].

Decision trees are tree structure classifiers. In a decision tree, a classification process is illustrated through the nodes and leaves of the tree. Nodes of a tree are classification rules, that perform thresholding or categorising on a feature of the feature set. Leaves of a tree assign classification labels to a test sample, that are classified by a series of classification rules defined on the nodes. Training a classification tree is to find an optimum tree structure according to information disparity measures. Using a decision tree to perform a classification can be found in brain segmentatio [66,67], blood vessel segmentation [68], diagnosis of breast cancer [69], and diagnosis of prostate cancer [70].

LDA is a linear classifier. It predicts the class of a sample through a linear combination of the features. It assumes the samples from different classes have a Gaussian distribution with the same variance but different means. Weights of the features and the interception of the linear function are usually estimated by using the maximum likelihood method. In [71], LDA is used in the segmentation of brain MRI images. Fraz et. al. used LDA to segment vasculature in retinal images [72]. Using LDA to diagnose pathological tissue observed in mammograms and MRI images is researched in [43–45].

The MLP classifier is one kind of artificial neural networks. It is composed of multiple layers. Each unit in the hidden layer is modelled with a sigmoid function, whose input is the output of the last layer. The network is trained by error back propagation and is able to tackle non-linear separable problems. Segmentation with the MLP is applied to localise masses in mammograms [73], cell nuclei in tissue sections of urine bladder tumours and brain tumours [74], and hard exudates in retinal images [75]. In addition, many researches are conducted to use the MLP in breast cancer diagnosis [76–78] and chromosomal aberration scoring [79].

The SVM classifier is a classification method that uses a decision plane to separate samples into different classes. The kernel trick is used in SVMs to deal with the non-linear separable problem. Applications of using the SVM for segmentation are aimed at retinal images [80, 81], MRI brain images [82] and mammograms [83]. Using the SVM for diagnosis is covered in many areas, such as lung computed tomography [46, 47], endoscopy [84, 85] and fundography [48].

In [49], the classification performances of KNN, decision trees and SVMs are com-

pared by testing on breast cancer grading, extent scoring of lymphocytic infiltration of breast cancer specimens, and prostate detection using MRI and histopathological images. It concludes that when there are sufficient training data, the performance of SVMs is the best.

Comparison among LDA, MLPs, and SVMs is conducted to diagnose benign and malign tumours using 130 digitised mammograms [86]. Similar researches can be found in dementia diagnosis and glaucoma diagnosis [87, 88], where the input features are qualitative descriptions of the diagnosis rules directly collected from patients. The highest accuracy of the SVM is shown through cross validation analysis. It is pointed out that the superiority results from the superior ability of generalising inner decision criteria in a high dimensional space [89].

We used SVMs in the erythema segmentation and scaling segmentation. Details of the SVM algorithm are described in Chapter 3 for the erythema segmentation. Due to the limited samples experimented in this thesis, we used KNN and decision trees for the severity scoring of psoriasis (see Chapter 5 for the details).

Graph based classifiers

A graph based classifier is a classification method that models the classification process by using the structure of the graph. One kind of widely used graph based classifiers is the Markov Random Field (MRF). When the MRF is applied to the image segmentation, the classification problem is converted into labelling vertices on a graph. The intensity distribution of the abnormality and the region homogeneity are modelled in the segmentation process. Due to this property, the MRF classifiers have more advantages in texture segmentation than the sample-based classifiers. The intensity distribution is often modelled as Gaussians in the MRF. Application of the Gaussian MRF can be found in the segmentation of MRI brain images [90], ultrasound images [91] and CT images [92]. In Chapter 4, the MRF modelling is described for the scaling segmentation.

2.4 Digital Image Processing in Assessment of Psoriasis Severity

Computer-aided diagnosis systems based on image processing have been researched to objectively measure the severity of psoriasis for more than ten years. Even though in the same period, dominant effort was spent on skin cancer diagnosis, more and more attention was devoted to psoriasis research. Image processing techniques are applied in psoriasis segmentation, severity assessment of erythema and scaling, and psoriasis changes analysis. The following is a review of these applications.

2.4.1 Psoriasis segmentation

Psoriasis segmentation is a preliminary step in the analysis of psoriasis severity. This part corresponds to the abnormality segmentation part of a computer-aided diagnosis system as illustrated in Figure 2.2. For psoriasis severity diagnosis, the regions of interest are psoriatic lesions, erythema, and scaling, since the three elements directly decide the severity degree of psoriasis. Related researches on segmentation of the three regions are reviewed below:

Psoriatic lesion segmentation

The segmentation method can be roughly divided into two kinds. One is a pixel-based method, and the other is a region-based method. The pixel-based method investigates colour spaces and only uses colour values of a single pixel for segmentation. The other is a region-based method. It not only uses colour information, but also considers the neighbouring information of the pixel. The region-based methods include texture-based classification, watershed and region-based active contours.

There are a number of pixel-based methods investigated for psoriasis segmentation. One of the mostly used approaches is thresholding. In [5], psoriatic lesions are segmented by thresholding G values in a normalised RGB colour space. However, this method mis-segments shadows as lesions. To resolve this problem, a local variable thresholding method is proposed in [6]. The threshold values change with local regions, which are small blocks in the image. The variable thresholds of local regions are decided based

on an assumption that grey-scale distribution of normal skin and psoriatic lesions are two Gaussians. This method removes shadings from psoriatic lesions. Nevertheless, this system is not accurate. Small psoriatic lesions are removed as shadings at the same time.

In [93, 94], the CIE $L^*a^*b^*$ colour space is used. The hue and chroma colour component are analysed in the colour space, and are used to separate psoriatic lesion pixels from normal skin pixels. The separation is based on the minimum Euclidean distance to the centroids of sampled psoriasis and normal skin pixels. Delgado et al. [7] extracted psoriatic lesions from skin images using quadratic discriminant analysis. This analysis assumes that the difference between the G and B values in the RGB colour space follows the Gaussian distribution. Nevertheless, shadow disturbance is still unresolved with these methods.

The region-based segmentation of psoriasis greatly reduces disturbance from uneven illuminance. In [8], fuzzy texture spectrum and colour value in a normalised RGB colour space are used as feature vectors for a neuro-fuzzy classifier to segment psoriasis. This method is improved in [9], where the algorithm computation complexity is decreased with an alternative classifier adapted from a multiresolution-decomposed signature subspace classifier. The drawback of these approaches is that they rely on the extraction of homogenous regions to model the training sets. The method of identifying homogeneous regions is reliable for large regions, but is less accurate in detecting small spot-shaped psoriatic lesions.

Fractal features are proposed to do the segmentation with the watershed algorithm, given the psoriatic lesion is with a more complex texture than the normal skin [95, 96]. However, when there are no great changes in the lesion, such as the guttate psoriasis, this method would be not suitable. In [97], geometric active contours are employed. After seed regions of psoriasis are localised, boundaries of the seeds gradually grows in the geometric active contour framework. [97] relies on initialisation of the seed regions, which are identified through a colour analysis. The problem is that this colour analysis only works for dull-red regions. When only scaling is contained in the lesion, this lesion would be mis-segmented as normal skin.

Erythema segmentation

So far, a few works have been conducted for the erythema segmentation. The performed research mainly focused on analysing the colour information of erythema in skin images. In [98], boundaries of erythema are identified through the subtraction operation between the baseline image and the image with erythema that is produced with the noxious stimulation of the baseline image. The minimum distance classifier is used in the CIE $L^*a^*b^*$ colour space to decide the erythema region. This research requires the baseline images, which are generally not available in the psoriasis treatment. In [99], erythema segmentation is performed using a region growing algorithm based on manually selected seeds. Neighbouring pixels with similar "a" values in the CIE $L^*a^*b^*$ colour space are aggregated to the growing region. [100] researches the erythema segmentation problem in the HSV colour space with a fuzzy c-means algorithm. Considering the calculation complexity, [100] just uses hue component (H) and the saturation component (S). However, in this research, only four cases are tested, which causes a deficiency.

Scaling segmentation

Scaling area is another sign for the evaluation of psoriasis severity. It is the result of an enhanced rate of epidermal cell production. Visually, it manifests itself from a few spots to a large area on inflamed skin. So far, the only research on scaliness segmentation has been implemented by Delgado et al. [7]. Scaliness is extracted from erythema using a clustering algorithm in a homogenous red region determined by a watershed algorithm. However, due to disturbance from uneven illuminance, human intervention is needed to determine the number of watersheds.

Nevertheless, all of the psoriatic lesion segmentation, erythema segmentation and scaling segmentation researches focus on plaque psoriasis [5–9, 93–97]. It is not applicable to apply these methods to a general psoriasis lesion, due to the various appearance changes of psoriatic lesions. In this thesis, psoriasis segmentation is decomposed into erythema segmentation and scaling segmentation, since a psoriatic lesion is either composed of erythema and scaling, or erythema alone, or scaling alone. Erythema is

segmented with an SVM by decomposing the skin into haemoglobin component and melanin component. Scaling is segmented through an MRF smoothing on an SVM classification result. A scaling contrast map is constructed and a Gabor texture is analysed in the feature analysis stage.

2.4.2 Treatment evaluation

The computer-aided image analysis for psoriasis treatment evaluation is based on segmentation of psoriatic lesions, erythema and scaling. In clinical practice, dermatologists use two ways to test the treatment efficacy. One method is tracking the progress of a single lesion over a period of time, and the other method is comparing the severity of different lesions. Thus, the researches on computer-aided psoriasis treatment efficacy evaluation are divided into two groups: the severity evaluation of different lesions and the treatment evaluation of a single lesion over a period. The severity evaluation of psoriasis corresponds to the diagnosis decision making part of a computer-aided diagnosis system (see Figure 2.2).

Severity evaluation of different lesions

As far as we know, there are a few works on computer-aided diagnosis of psoriasis severity. The conducted works evaluate the psoriasis severity from the aspects of erythema severity and scaling severity.

In [7], the erythema severity of psoriasis is evaluated by using the mean colour value of the lesion in the RGB colour space. The mean values are then used to grade the severity of lesions using a KNN classification. In [101], the a^* component in the $L^*a^*b^*$ colour space is considered for the erythema severity scoring. The erythema severity is measured by using the difference between the means of the a^* component values of the lesion and skin. [7] and [101] take into account of the colour variation in a lesion, and are reliable for the measurement of a large part of skin is affected by psoriasis, compared to the measurement performed by colorimeters and spectrometers.

To the best of our knowledge the only work that attempts to grade the severity of

scaling is given in [7], where the severity of scaling is derived by building a decision tree using the area of scaling inside the lesion. There are 46 psoriasis images collected to achieve a maximum possible diversity. It is reported that the proposed method is able to correct misdiagnosis made by dermatologists.

In this thesis, the performance of the severity scoring using 2D skin images is improved. The haemoglobin and melanin components are used to describe the severity of erythema. The Gabor feature and the scaling area are applied to describe the severity of scaling. Linear correlations between the severities and the proposed severity features are analysed. Additionally, the erythema severity and scaling severity are graded using a KNN classification and a decision tree respectively. An experiment shows that the correlations of the proposed severity descriptors and the classification results are better than previous methods.

Treatment evaluation of a single lesion over a period

The evolution of a single lesion is often evaluated from the change of psoriatic lesions in the image. In [14, 102] it is assumed that psoriatic lesions change their shape and size slowly over time, and the significant change only happens within the lesions. Images of the same lesion, which are photographed at different time points, are registered at first based on lesion segmentation results. Secondly corresponding portions of the lesion are compared between the images.

Delgado et al. [10] compared a series of aligned images with change detection technologies used in geo-informatics. The images are taken using a specifically designed imaging system to ensure that the images are photographed under the same illumination and the influence of shadows and specular reflection is minimised [10]. Changes are analysed using image subtraction, PCA, and Multivariate Alteration Detection (MAD). In addition, the degree of the severity changes is classified with a decision tree.

However, the chronic psoriasis treatment often spans more than a month. In the long-term treatment, psoriatic lesions do not only change within their boundaries, but also the boundary itself changes. Thus change analysis through image registration of the lesions is not suitable for comparisons in a chronic treatment.

In this thesis, changes of psoriatic lesions are assessed based on the changes of severity features. The consistency between the features for assessing changes in severity and the PASI scores assigned by clinicians is validated using a multiple linear regression analysis.

2.5 Statement of problems solved in this thesis

There are three major gaps between existing methods and objectively diagnosing the efficacy of psoriasis treatment through skin images:

- Existing psoriasis segmentation methods focus on segmentation of plaque psoriasis only. These methods are not available for the segmentation of a general psoriatic lesion, since appearances of psoriatic lesions are quite different for different types of psoriasis.
- Existing psoriasis severity assessment methods just examine a few kinds of severity features. There is still a space to improve the accuracy of the severity assessment with new severity features.
- Existing psoriasis change assessment methods rely on the registration of psoriatic lesions. They are not available for the efficacy assessment of long term treatments, during which lesion boundaries and contents can dramatically change.

In this chapter, psoriasis and its severity assessment are introduced. Currently used psoriasis severity assessment methods bear unavoidable inter- and intra- observer variances. Though several devices are investigated to provide objective and quantitative measurements, manually operating these devices is tedious and prevents wide usage. Since cameras are commonly used in psoriasis treatment to record psoriasis appearance and the severity accordingly, using cameras to assess the severity of psoriasis automatically and reliably receive increasing interest.

This thesis develops a computer-aided image analysis system to reliably evaluate the psoriasis severity. There are two parts in the system, psoriasis segmentation and psoriasis treatment efficacy evaluation. In the psoriasis segmentation stage, this thesis proposes a segmentation strategy for general psoriatic lesions. The psoriasis segmentation method

proposed in this thesis is divided into two parts, erythema segmentation and scaling segmentation. In the psoriasis treatment efficacy evaluation stage, the treatment efficacy is evaluated from the aspects of psoriasis severity and changes of psoriasis severity during the treatment.

In Chapter 3, erythema is segmented from normal skin by using ICA to decompose a skin colour into its haemoglobin and melanin components. An SVM classifier is used to perform a pixel-based segmentation. The proposed segmentation method has a good sensitivity to identify erythema pixels, especially for mild erythema.

In Chapter 4, scaling segmentation is performed through a pixel labelling method. In the feature analysis stage, a scaling contrast map is constructed to enhance the contrast between scaling and its surrounding erythema, and a Gabor texture is analysed to differentiate rough scaling from smooth normal skin. This is followed by an unsupervised classification method to identify scaling pixels in psoriasis skin images. Training sets are collected automatically for each image using a soft constrained K means, and an SVM smoothed by an MRF is proposed to perform the segmentation. This method is tested on images with different skin colours and photographed in different imaging situations. The robustness of the scaling segmentation method is demonstrated.

In Chapter 5, psoriasis severity is assessed from erythema severity and scaling severity. The erythema severity is assessed by using the haemoglobin and melanin components proposed in the erythema segmentation stage. A KNN classifier is used to perform the severity classification. The scaling severity is assessed by using the relative scaling area and a Gabor texture that is proposed as a roughness feature in the scaling segmentation. A decision tree is performed to estimate scaling severity scores. The reliability and the superiority of the proposed methods are validated by comparing with previously investigated severity features and scoring classifiers.

In Chapter 6, psoriasis evolution is evaluated by using the changes of the erythema severity and scaling severity. This method avoids registration of lesion images, and is applicable to the assessment of treatment efficacy in long-term treatments. Building on the severity features proposed for the severity assessment, severity change features are proposed in this chapter, and are validated through a multiple linear regression analysis.

By using the multiple linear regression analysis, it is possible to assess the small severity changes that cannot be detected by PASI alone.

Part III

Psoriatic lesion segmentation

Chapter 3
Erythema Segmentation

3.1 Introduction

ERYTHEMA appears predominantly at the initial stages of psoriasis. Some psoriatic lesions may consist of erythema alone without any scaling. The diagnosis of erythema is according to its colour. Since erythema is caused by a skin inflammation, it shows a much redder colour than the normal skin. Thus, the erythema segmentation algorithm proposed in this chapter applies a skin colouring model to perform the segmentation.

The erythema segmentation algorithm is performed in a feature space closely related with the skin colouring model. Specifically, this space is based on the decomposition of a skin colour into the independent haemoglobin component and melanin component. By using the colour features extracted through the skin colour decomposition, a pixel-based classification method is proposed to separate erythema pixels from non-erythema pixels in the psoriasis skin images.

The algorithm is composed of three steps: (1) separating skin from background, (2) skin colour decomposition, and (3) erythema pixel extraction. In Section 3.2, skin is separated from background using a histogram-based Bayesian classifier. We separated skin from background to facilitate the following steps to focus on the skin for psoriasis images containing background objects, such as clothes and wallpapers. In Section 3.3, based on a skin imaging model, the skin colour is decomposed into melanin and haemoglobin components using an Independent Component Analysis (ICA). In Section 3.4, a Support Vector Machine (SVM) is used to separate erythema pixels from non-erythema pixels us-

ing the skin decomposition results. The parameter validation for the skin decomposition and experimental results for erythema segmentation are presented in Section 3.5.

3.2 Separating skin from background

A number of skin segmentation methods have been proposed for face recognition, human tracking and gesture analysis. The skin segmentation is performed by using skin colours, since skin colour is robust information, and is not affected by the position of people and image scaling. Explicit skin-colour space thresholding, Bayesian classifiers and Gaussian classifiers are three kinds of most popular methods for skin segmentation [103]. The explicit skin-colour space thresholding method defines boundaries of human skin in a certain colour space. The boundaries are often given by fixed multiple colour ranges [104]. The Bayesian classifier works together with colour histograms, which gives a probability distribution for skin colours. The Bayesian rule is used for distinguishing between skin pixels and non-skin pixels [105]. The Gaussian classifier approach assumes that the distribution of colours in the skin pixels and the distribution of colours in the background pixels are two different Gaussians. Comparison of probabilities of colour values in the Gaussians is performed [106]. Among the methods, the histogram-based Bayesian classifier performs best. The reason is that the colour histogram is a stable object representation unaffected by occlusion, while the explicit skin-colour space thresholding is easily biased by illumination and the threshold values are hard to identify [103]. Additionally, the histogram-based Bayesian classifier is computationally faster than the Gaussian models [107].

This section presents a method of segmenting body skin using a histogram-based Bayesian classifier. Skin pixels are at first classified through a Bayesian modelling of skin colour in the YCbCr colour space, since in the YCbCr colour space, the skin colour is compactly clustered. Two kinds of one dimensional histograms are examined. One is based on the **Cb** colour component and the other is based on the **Cr** component. Secondly, misclassified pixels are corrected by using a neighbourhood connectivity analysis.

3.2.1 A Bayesian modelling of skin colour in the YCbCr colour space

The YCbCr colour space

In a colour space, an arbitrary colour is represented as a combination of colour components, where each colour component is projected onto one of the coordinate axes. In this section, skin colour modelling is implemented in the YCbCr colour space to differentiate skin from background.

In the YCbCr colour space, a colour is described by the aspects of luminance (**Y**) and chrominance (**Cb** and **Cr**). The **Cb** component measures the difference between the blue and a reference value and the **Cr** component measures the difference between the red and a reference value.

The YCbCr colour space is different from the widely used RGB colour space, where the **R**, **G**, and **B** represent the three primary colours: red, green and blue respectively. The RGB colour information in a colour image is directly outputted by a digital camera. The **R**, **G**, and **B** values are proportional to the amounts of the primary lights needed to be emitted to form a given colour [108]. The RGB colour space is not preferable for skin detection, since there is a high colour correlation among the **R**, **G**, and **B** values [105].

The YCbCr colour space is linearly transformed from the RGB colour space and is orthogonal. It reduces the redundancy that exists in the RGB colour space, and represents colour as independently as possible [103]. We use the YCbCr colour space here, because the skin colour is compactly clustered in the YCbCr colour space [109]. Thus, the YCbCr colour space is well suited to skin detection [109–111].

The linear transformation from the RGB colour space to the YCbCr colour space is given by:

$$\begin{aligned} \mathbf{Y} &= 0.299\mathbf{R} + 0.587\mathbf{G} + 0.114\mathbf{B} \\ \mathbf{Cb} &= -0.172\mathbf{R} - 0.339\mathbf{G} + 0.511\mathbf{B} + 128 \\ \mathbf{Cr} &= 0.511\mathbf{R} - 0.428\mathbf{G} - 0.083\mathbf{B} + 128 \end{aligned} \tag{3.1}$$

where the **R**, **G**, and **B** are 8-bit integers with the range $[0, 255]$. The luminance **Y** value is

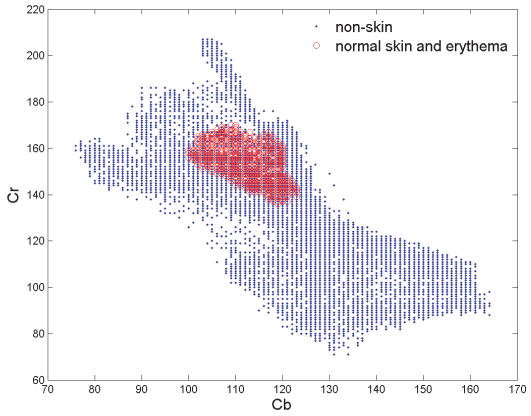


Figure 3.1: Distribution of skin colour compared to non-skin colour in CbCr space.

in the range $[16, 235]$, and the chrominance **Cb** and **Cr** values are in the range $[16, 240]$.

In this work, the luminance component **Y** is discarded to reduce the disturbance from uneven illumination, e.g. highlights and shadows. Pixels labelled as skin as well as erythema are collected from the images in the database. It is found that even though erythema is redder than normal skin, colour values of skin, including normal skin and erythema, still cluster in a small region compared with non-skin pixels in the **Cb**-**Cr** component space (as shown in Figure 3.1).

The Bayesian modelling of skin colour

Bayesian modelling separates skin pixels from non-skin pixels according to the Bayesian rule. Given a colour component, the likelihood of different classes are compared, and the feature is assigned with a class corresponding to the biggest likelihood. It is a simple learning based classifier.

When the Bayesian classifier is applied to the skin segmentation, the segmentation problem is treated as a binary classification problem. One class is the skin pixels, and the other is for the pixels in the background (called as non-skin pixels here). Probability distributions of skin tones and non-skin tones are represented with normalised colour histograms, which are defined as discrete functions parameterised with a colour intensity:

$$P(x) = \frac{c(x)}{c_t} \tag{3.2}$$

where x is the xth intensity level of a colour component, $c(x)$ is the total number of pixels with intensity level x, and c_t is the total number of pixels in the histogram. The normalised histogram $P(x)$ shows the probability of the occurrence of the intensity level x in the histogram.

Given skin and non-skin histograms, which are obtained from a group of skin pixels and non-skin pixels respectively collected in a training set, the conditional probability of skin pixels $P(x|\text{skin})$ (given a skin pixel, the probability of a pixel colour x being skin) and non-skin pixels $P(x|-\text{skin})$ (given a non-skin pixel, the probability of a pixel colour x being non-skin) are represented as follows:

$$\begin{aligned} P(x|\text{skin}) &= \frac{c_s(x)}{c_s} \\ P(x|-\text{skin}) &= \frac{c_n(x)}{c_n} \end{aligned} \tag{3.3}$$

Where $c_s(x)$ and $c_n(x)$ are respectively the number of the skin pixels and non-skin pixels with the colour x. c_s and c_n represent the total number of skin pixels and non-skin pixels respectively in the histograms.

A skin tone classifier is built using a Bayesian maximum likelihood estimation. Given a colour value x, the likelihood of being skin $P(\text{skin}|\text{x})$ and non-skin $P(-\text{skin}|\text{x})$ is defined as:

$$\begin{aligned} P(\text{skin}|\text{x}) &= \frac{P(x|\text{skin})\text{P}(\text{skin})}{P(x|\text{skin})\text{P}(\text{skin}) + \text{P}(\text{x}|-\text{skin})\text{P}(-\text{skin})} \\ P(-\text{skin}|\text{x}) &= \frac{P(x|-\text{skin})\text{P}(-\text{skin})}{P(x|\text{skin})\text{P}(\text{skin}) + \text{P}(\text{x}|-\text{skin})\text{P}(-\text{skin})} \end{aligned} \tag{3.4}$$

A comparison of likelihoods $P(\text{skin}|\text{x})$ and $P(-\text{skin}|\text{x})$ can be performed through the ratio of $P(\text{skin}|\text{x})$ to $P(-\text{skin}|\text{x})$:

$$\frac{P(\text{skin}|\text{x})}{P(-\text{skin}|\text{x})} = \frac{P(x|\text{skin})\text{P}(\text{skin})}{P(x|-\text{skin})\text{P}(-\text{skin})} \tag{3.5}$$

If a pixel colour x is such that

$$\frac{P_{\text{skin}}(x|\text{skin})}{P_{-\text{skin}}(x|-\text{skin})} \geq t_{\text{skin}} \tag{3.6}$$

exceeding the threshold t_{skin}, then x is labeled as a potential skin pixel. The threshold t_{skin} is a function of

$$t_{\text{skin}} = \frac{c_p P(-\text{skin})}{c_n P(\text{skin})} \tag{3.7}$$

where c_p is the introduced cost of a false positive, and c_n is the introduced cost of a false negative. $P(-\text{skin})$ and $P(\text{skin})$ are prior probabilities of skin and non-skin respectively, which are case dependent. Thus, comparisons of the likelihood are converted to comparisons of the conditional probabilities.

Implementation of the Bayesian skin classifier

In this work, two Bayesian maximum likelihoods are estimated: one is parameterised with colour intensities of the **Cb** component, and the other is parameterised with the **Cr** component. Trading off between true positives and false positives, the threshold t_{skin} is set to be 0.2 for the **Cb** component and 0.1 for the **Cr** component according to the experiment. The sets of skin colour intensities in the **Cb** component $X_{\text{skin}}^{\text{Cb}}$ and the **Cr** component $X_{\text{skin}}^{\text{Cr}}$ are respectively defined as:

$$\begin{aligned} X_{\text{skin}}^{\text{Cb}} &= \{x_{\text{Cb}} | \frac{P_{\text{skin}}(x_{\text{Cb}}|\text{skin})}{P_{\text{skin}}(x_{\text{Cb}}|-\text{skin})} \geq 0.2\} \\ X_{\text{skin}}^{\text{Cr}} &= \{x_{\text{Cr}} | \frac{P_{\text{skin}}(x_{\text{Cr}}|\text{skin})}{P_{\text{skin}}(x_{\text{Cr}}|-\text{skin})} \geq 0.1\} \end{aligned} \tag{3.8}$$

where x_{Cb} and x_{Cr} are the colour intensity in the **Cb** component and the **Cr** component respectively.

Let A_{skin}^{Cb} be the pixel set with colour intensities of the **Cb** component in the set X_{skin}^{Cb}, and A_{skin}^{Cr} be the pixel set with colour intensities of the **Cr** component in the set X_{skin}^{Cr}. The classification result of the Bayesian maximum likelihood classifier $A_{skin}^{'}$ is the intersection of A_{skin}^{Cb} and A_{skin}^{Cr} :

$$A_{skin}^{'} = A_{skin}^{Cb} \bigcap A_{skin}^{Cr} \tag{3.9}$$

The classification result is presented using a binary image, where pixels belonging to the set $A_{skin}^{'}$ are marked with '1', and the rest pixels are marked with '0'.

Here, the **Cb** and **Cr** components are considered separately rather than considering them jointly in a 2D colour space. Considering each component separately allows us to loosen the constraints on the two dimensions. It also allows us to determine the thresholds for each dimension separately which is simpler than using a 2D histogram.

3.2.2 Connectivity analysis for the correction of misclassified skin pixels

The problem with the classification above is that there are regions within the skin and background that could be misclassified by the Bayesian classifier. In the binary image that results from the Bayesian classifier, small holes, that are marked with '0' and shown in black in the white coloured skin regions, are the result of extreme redness and scaliness of abnormal skin and shading. They are misclassified skin regions. Patches, that are marked with '1' and are shown in white on the black coloured background, are usually due to the colour similarity with skin. They are misclassified background regions.

The analysis of the connectivity properties of regions leads us to conclude that there are two kinds of misclassified regions, both of which can be detected using an 8-adjacent connectivity analysis. The adjacent connectivity examines the connection between two pixels. When the two pixels are 8-adjacent, they are connected side by side, or they share the same corner. Mathematically, if two pixels points: p and q, are 8-adjacent, then:

$$\begin{aligned} N_8(p,q) &= \max\{|x_p - x_q|, |y_p - y_q|\} \\ &= 1 \end{aligned} \tag{3.10}$$

where (x_p, y_p) is the coordinate of pixel p, and (x_q, y_q) is the coordinate of pixel q.

A region is defined to be a set of pixels, such that for any two pixels in the region, there is a sequence of 8-adjacent points that connects them. Specifically, pixels in a hole region are 8-adjacent connected and have values '0', while pixels in a patch region are 8-adjacent connected and have values '1'. A labelling strategy outlined in [112] is used to do the connectivity analysis. In [112], different connected regions are labelled with different labels. The labelling strategy scans pixels having a target value: '0' or '1' in the binary image. It assigns a new label to a currently scanned pixel, when its neighbourhood pixels are not labelled. Otherwise, the currently scanned pixel is assigned the same label as the neighbourhood pixels.

By means of counting pixels in the connected regions, the areas of the connected regions are measured to identify whether the regions are misclassified. Since the hole regions caused by abnormal skin and shadows are small, if the number of pixels in a hole region is less than a threshold value, pixels in the hole region are reclassified as skin. In this work, the threshold value is experimentally set as 0.008 times of the current image size. A skin region is always the largest patch in the image, when there are multiple patches in the image. A patch region where the number of pixels is less than the patch with the maximum number of pixels is a non-skin region. These non-maximal patch regions are reclassified as background, for example, this is how the scarf region in Figure 3.2 was eliminated.

Figure 3.2 shows two examples of the skin segmentation. The Bayesian classification results are shown as binary images in the second column of Figure 3.2. Shadowed skin in the first original image is misclassified as background by the Bayesian classifier. Background patches with similar skin colour in the second original image are misclassified as skin. In the third column of Figure 3.2, the misclassifications are eliminated in the final skin segmentation results through the connectivity analysis examination.

(a) (b) (c)

Figure 3.2: Skin segmentation result. (a) Original image; (b) Bayesian classifier segmentation result; (c) Final skin segmentation result.

3.3 Skin colour decomposition

After identification of the skin region, the skin colour is decomposed into a melanin component and a haemoglobin component. These two components correlate directly with skin colouring and can be used to distinguish between erythema and other non-inflamed skin such as normal skin and moles. In this thesis, the purpose of the skin decomposition is to derive a feature space that distinctly differentiates erythema from non-erythema skin.

Figure 3.3 shows a dichromatic reflection model of skin. Human skin is a turbid media with a multi-layered structure. Except for the 6-8% incident light reflected directly on skin surface, most light enters the skin, and is absorbed by the melanin and haemoglobin pigments. The melanin pigment is from the epidermal layer and causes skin to present brown or black; the haemoglobin pigment is from blood cells in the dermal layer and brings reddish to skin colour. The quantities of the two pigments are different from individual to individual, and this causes the variety of skin colours that we see. When it comes to erythema, its accompanying skin inflammation causes an increase of haemoglobin pigment and the abnormal redness shown on the skin surface. The normal skin maintains a normal mixture of the haemoglobin pigment and melanin pigment.

Additionally, the skin colour in the digitised image is a reflection of the skin pigments. As shown in Figure 3.3, the incident light passes through the epidermis layer and the

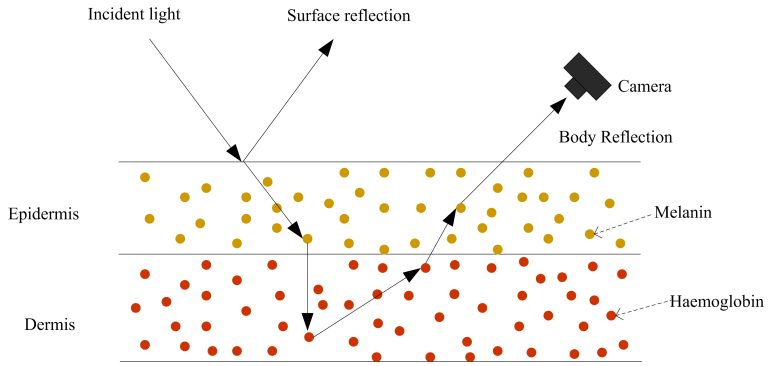


Figure 3.3: Dichromatic reflection model of skin.

dermis layer, and is absorbed by the melanin and haemoglobin pigments. It is emitted from the skin as body reflection. When the body reflection enters a camera, the skin image is photographed and the skin colour is shown in the image. Therefore, the distribution of skin pigments is reflected on the skin image.

It is assumed that the haemoglobin and melanin pigments are decided by the and melanin components respectively and the two components are independent to each other. In this work ICA is used to decompose a skin colour into the two pigment components and to extract the quantities of a skin colour associated with the two pigment components. This is an extension of the method proposed in [113], that synthesises skin colour by using a linear skin colouring model factoring in melanin and haemoglobin components extracted by ICA. There are four steps in the skin colour decomposition algorithm:

1. A skin colour model is defined in the three dimensional optical density domain, so that the skin colour density can be linearly represented by the melanin and haemoglobin components;
2. Principle Component Analysis (PCA) is used to reduce the dimension of skin colours to two dimensions. In the two dimensional space, the major skin colour information is preserved, and the redundant information caused by other pigments and skin structure is discarded;
3. Working on the two dimensional space, ICA is applied to find out the independent components: the melanin and haemoglobin components; and
4. The melanin and haemoglobin quantities of a skin colour are derived by using the

skin colour model and the extracted melanin and haemoglobin components.

3.3.1 Skin colour modelling in the optical density domain

According to the Beer-Lamber law [114], the ratio of the intensity of the observed light to the intensity of the incident light is in a negative logarithmic relationship with the product of light absorption and the distance that the light travels. Since the intensity of the incident light and the distance of the light path can be assumed to be constant, the linearity between the observed colour and the pure colour intensity is assumed to be in the optical density domain. The colour density vector in the optical density domain is derived through a transformation from the RGB colour space using the negative logarithm operation:

$$\mathbf{l}_{x,y} = [-\log(\mathrm{r_{x,y}}), -\log(\mathrm{g_{x,y}}), -\log(\mathrm{b_{x,y}})]^{\mathrm{t}} \tag{3.11}$$

Where $r_{x,y}$, $g_{x,y}$ and $b_{x,y}$ are normalised colour densities of a pixel at coordinate (x, y) for the **R**, **G**and **B** component of the RGB colour space respectively, the ranges of $r_{x,y}$, $g_{x,y}$ and $b_{x,y}$ are $[0, 1]$, and the superscript t is the transposition operation.

Based on the assumption that the spatial variations of skin colours are caused by the melanin and haemoglobin components and their quantities are mutually independent, the colour density vector is modelled as:

$$\begin{aligned} \mathbf{l}_{x,y} &= \mathbf{c}^m q^m_{x,y} + \mathbf{c}^h q^h_{x,y} + \Delta \\ &= C\mathbf{q}_{x,y} + \Delta \end{aligned} \tag{3.12}$$

where $\mathbf{c}^m$ and $\mathbf{c}^h$ are the melanin and haemoglobin components, $q^m_{x,y}$ and $q^h_{x,y}$ are relative quantities of the melanin and haemoglobin components respectively, and Δ is a spatially stationary vector caused by other pigments and skin structure. The equation can be rewritten using a pure density matrix $C = [\mathbf{c}^m, \mathbf{c}^h]$ and quantity vector $\mathbf{q}_{x,y} = [q^m_{x,y}, q^h_{x,y}]$. Figure 3.4 shows this model.

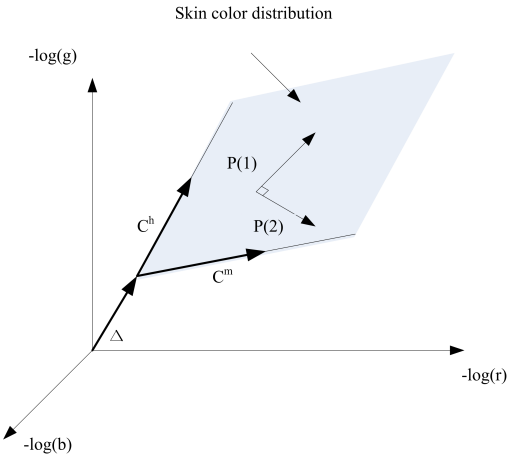


Figure 3.4: Skin colour model in optical density domain.

3.3.2 Reducing the dimension of skin colours by using PCA

PCA is a useful dimensionality reduction method for removing redundant information. It performs an eigen-decomposition to identify mutually uncorrelated feature vectors. The amount of information represented by the feature vector is related with the corresponding eigenvalue. The vector corresponding to the biggest eigenvalue represents the biggest variance and is the most informative.

PCA is used in this thesis to estimate the information contributed by the melanin and haemoglobin components to the major skin colours, and to suppress other colours found in the skin, such as shadow colours caused by skin structures. Thus, the two most informative vectors in skin colours are required to identify in the PCA process. The resulting two principle components, $\mathbf{p}(1)$ and $\mathbf{p}(2)$, are shown in Figure 3.4. The space spanned by $\mathbf{p}(1)$ and $\mathbf{p}(2)$ is the same as the space spanned by the melanin and haemoglobin components, $\mathbf{c}^m$ and $\mathbf{c}^h$, that is where the skin colour is distributed. The procedure of PCA is introduced below:

1. Let L be a three-row matrix composed of 3-element colour intensities of the image in the spectral density domain.

$$L = [\mathbf{l}_{1,1}, ..., \mathbf{l}_{x,y}, ..., \mathbf{l}_{m,n}] \tag{3.13}$$

where m and n are the width and length of the image respectively.

2. The PCA process calculates the covariance of the normalised L, and conducts an eigen-decomposition of the covariance matrix, given by:

$$P\Lambda P^t = \frac{1}{n \times m}(L - \bar{\mathbf{l}})(L - \bar{\mathbf{l}})^t \tag{3.14}$$

where on the right side of the equation is the covariance matrix: $\bar{\mathbf{l}}$ is the mean vector of L and $n \times m$ is the number of columns in L.

$$\bar{\mathbf{l}} = \frac{1}{n \times m} \sum_{x=1}^{x=m} \sum_{y=1}^{y=n} \mathbf{l}_{x,y} \tag{3.15}$$

On the left hand side of Eq. (3.14), Λ is a diagonal matrix:

$$\Lambda = \text{diag}[\lambda(1), \lambda(2), \lambda(3)] \tag{3.16}$$

where the diagonal elements, $\lambda(1)$, $\lambda(2)$, and $\lambda(3)$, are the eigenvalues of the covariance in descending order, and P is a 3×3 matrix composed of eigenvectors: $\mathbf{p}(1)$, $\mathbf{p}(2)$ and $\mathbf{p}(3)$:

$$P = [\mathbf{p}(1), \mathbf{p}(2), \mathbf{p}(3)] \tag{3.17}$$

The order of the eigenvectors in the matrix P corresponds to the order of the eigenvalues in the diagonal matrix Λ.

3. The principle component vectors $\mathbf{p}(1)$ and $\mathbf{p}(2)$ are chosen, since they correspond to the first two biggest eigenvalues, and deliver the majority information of the skin colour.

After identification of the principle component vectors $\mathbf{p}(1)$ and $\mathbf{p}(2)$, the colour density vector of a pixel is re-expressed as:

$$\mathbf{l}_{x,y} = \tilde{P}\tilde{P}^t\mathbf{l}_{x,y} + (I - \tilde{P}\tilde{P}^t)\mathbf{l}_{x,y} \tag{3.18}$$

where I is an identity matrix, and $\tilde{P}\tilde{P}^t$ is the projection matrix$[\mathbf{p}(1), \mathbf{p}(2)][\mathbf{p}(1), \mathbf{p}(2)]^t$.

The first summand of Eq. (3.18) projects the colour density vector to the 2D plane spanned by $\mathbf{p}(1)$ and $\mathbf{p}(2)$. The second summand corresponds to the colour intensity in the one dimensional space spanned by $\mathbf{p}(3)$.

Transforming the skin colour $\mathbf{l}_{x,y}$ from the optical density domain to the 2D space spanned by basis vectors: $\mathbf{p}(1)$ and $\mathbf{p}(2)$, we have the corresponding value $\mathbf{w}_{x,y}$ associated with the basis:

$$\begin{aligned} \mathbf{w}_{x,y} &= \tilde{P}^t \mathbf{l}_{x,y} \\ &= \tilde{P}^t (C\mathbf{q}_{x,y} + \Delta) \\ &= \tilde{P}^t C \mathbf{q}'_{x,y} \end{aligned} \tag{3.19}$$

where $\mathbf{q}'_{x,y} = \mathbf{q}_{x,y} + (\tilde{P}^t C)^{-1} \tilde{P}^t \Delta$.

3.3.3 Extracting the haemoglobin and melanin components by using ICA

Subsequently, considering that $\mathbf{w}_{x,y}$ is a mixture of the melanin component $\mathbf{c}^m$ and the haemoglobin component $\mathbf{c}^h$, ICA is applied to estimate the melanin component $\mathbf{c}^m$ and the haemoglobin component $\mathbf{c}^h$.

ICA is a blind source separation method. It recovers source signals by assuming that the mixed signal is generated by a linear combination of the source signals that are independent to each other. Source signals are recovered by measurement of non-Gaussianity, since according to the central limit theorem, the probability distribution of a source signal is less gaussian than distribution of the sum of independent source signals. For an easy computation, whitening a signal to have unit variance and zero mean is required in the preprocess.

In this work, the whitened value $\mathbf{e}_{x,y}$ with unit variance and zero mean is given by:

$$\begin{aligned} \mathbf{e}_{x,y} &= \tilde{\Lambda}^{-\frac{1}{2}} [\mathbf{w}_{x,y} - \bar{\mathbf{w}}] \\ &= \tilde{\Lambda}^{-\frac{1}{2}} \tilde{P}^t C [\mathbf{q}'_{x,y} - (\tilde{P}^t C)^{-1} \bar{\mathbf{w}}] \\ &= H \mathbf{s}_{x,y} \end{aligned} \tag{3.20}$$

where $\tilde{\Lambda} = \text{diag}[\lambda(1), \lambda(2)]$ and $\bar{\mathbf{w}}$ is the mean vector, $\bar{\mathbf{w}} = \frac{1}{n \times m} \sum_{x=1}^{x=m} \sum_{y=1}^{y=n} \mathbf{w}_{x,y}$. The whitened value $\mathbf{e}_{x,y}$ is re-expressed using a separation matrix $H = \tilde{\Lambda}^{-\frac{1}{2}} P^t C$ and the vector $\mathbf{s}_{x,y} = [s_{x,y}(1), s_{x,y}(2)]^t = \mathbf{q}'_{x,y} - (\tilde{P}^t C)^{-1} \bar{\mathbf{w}}$, where $s_{x,y}(1)$ and $s_{x,y}(2)$ are independent to each other, and are source signals that can be recovered by ICA. In the following, we briefly introduce the ICA algorithm.

ICA measures the non-Gaussianity with negentropy in this work. The negentropy of a random variable is defined as the difference of the entropy of a Gaussian variable and the entropy of the random variable. Since a variable with a Gaussian distribution has a maximum entropy, the negentropy is larger, when the random variable is less Gaussian. For a small computational cost, a negentropy approximation is used to approximately describe the negentropy. A source signal can be identified by maximising the negentropy approximation $J(\cdot)$:

$$\begin{aligned} &\text{maximise} \quad J(\mathbf{a}) = [E\{G(\mathbf{a}^t \mathbf{e})\} - E\{G(v)\}]^2 \\ &\text{subject to} \quad E\{(\mathbf{a}^t \mathbf{e})^2\} = 1 \end{aligned} \tag{3.21}$$

where $\mathbf{a}$ is a two dimensional vector, $\mathbf{e}$ is a variable with samples $\mathbf{e}_{x,y}$ ((x, y) goes through all the coordinates in the image), v is a standardised Gaussian variable, E represents the expectation, and $G(\cdot)$ is a nonquadratic function. Considering simplicity and efficiency, $G(x)$ is set to be $\frac{1}{4}x^4$ in this work. This maximisation problem can be solved using the fixed-point iteration scheme proposed in [115]. The other source signal can be estimated using the same method with a hierarchical decorrelation. Thus, we have:

$$s_{x,y}(1) = \mathbf{a}_1^t \mathbf{e}_{x,y} \tag{3.22}$$

$$s_{x,y}(2) = \mathbf{a}_2^t \mathbf{e}_{x,y} \tag{3.23}$$

where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the estimated two dimensional vectors for $s_{x,y}(1)$ and $s_{x,y}(2)$ respectively.

Let $A = [\mathbf{a}_1, \mathbf{a}_1]^t$, we have:

$$H = A^{-1} \tag{3.24}$$

According to $H = \tilde{\Lambda}^{-\frac{1}{2}} P^t C$, the estimated pure density matrix $\tilde{C}$ is:

$$\tilde{C} = (\tilde{\Lambda}^{-\frac{1}{2}} P^t)^{-1} H \tag{3.25}$$

Substituting Eq. (3.25) into Eq. (3.12), the estimated quantity vector of pigments $\tilde{\mathbf{q}}_{x,y}$ is:

$$\tilde{\mathbf{q}}_{x,y} = \tilde{C}^{-1} \mathbf{l}_{x,y} - \mathbf{d} \tag{3.26}$$

where $\mathbf{d}$ is defined by $\tilde{C}^{-1}\Delta$. Since Δ is unknown, then assuming that the smallest value of $\tilde{\mathbf{q}}_{x,y}$ is zero, we have:

$$\mathbf{d} = \min_{x,y}(\tilde{C}^{-1} \mathbf{l}_{x,y}) \tag{3.27}$$

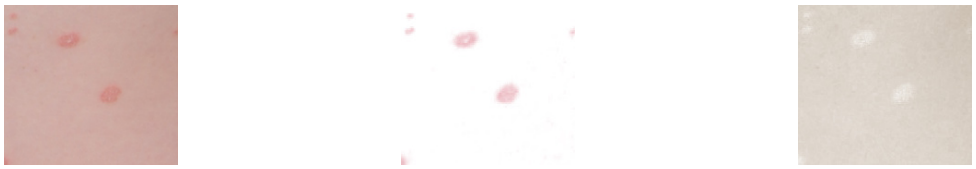
3.3.4 Deriving haemoglobin and melanin quantities

A skin colour is then decomposed as:

$$\tilde{\mathbf{I}}_{x,y} = \tilde{C}(K\tilde{\mathbf{q}}^t_{x,y} + j\mathbf{d}) + j\Delta \tag{3.28}$$

where $\tilde{\mathbf{I}}_{x,y}$ is the synthesised skin colour, $\tilde{C}$ is the estimated $[\mathbf{c}^m, \mathbf{c}^h]$, and K and j are introduced synthesis parameters. In this thesis, the spatial effect is removed with $j = 0$. Setting $K = \text{diag}[1, 0]$ and $K = \text{diag}[0, 1]$, the colour intensity of the melanin component $\tilde{\mathbf{I}}^m_{x,y}$ and the colour intensity of the haemoglobin component $\tilde{\mathbf{I}}^h_{x,y}$ are obtained in the optical density domain.

Transforming $\tilde{\mathbf{I}}^m_{x,y}$ and $\tilde{\mathbf{I}}^h_{x,y}$ from the optical density domain into the RGB colour space, we have:



(a) (b) (c)

Figure 3.5: ICA decomposition of a skin image. (a) Original skin image; (b) Haemoglobin skin image; (c) Melanin skin image.

$$\tilde{I}^{m}_{x,y} = \exp(-\tilde{\mathrm{I}}^{\mathrm{m}}_{\mathrm{x,y}}) \qquad \tilde{\mathrm{I}}^{\mathrm{h}}_{\mathrm{x,y}} = \exp(-\tilde{\mathrm{I}}^{\mathrm{h}}_{\mathrm{x,y}}) \tag{3.29}$$

where $\tilde{I}^{m}_{x,y}$ and $\tilde{I}^{h}_{x,y}$ are the estimated colour intensities of the melanin component and the haemoglobin component in the RGB colour space respectively.

Thus, the haemoglobin skin image and the melanin skin image are extracted in the RGB colour space. The haemoglobin skin image only reflects the pure colour intensities associated with the haemoglobin component, and in the melanin skin image, only the pure colour intensities of the melanin component are shown. Figure 3.5 shows an example of the ICA decomposition. In Figure 3.5 (a), erythema is shown as the red part in the original skin image. Due to its high association with the haemoglobin component, it presents as a much redder part than the surrounding normal skin in the haemoglobin skin image Figure 3.5 (b). In the melanin image Figure 3.5 (c), the erythema is presents as a lighter brown part.

3.4 Erythema pixel extraction

After the skin decomposition, erythema and non-erythema skin are represented with haemoglobin and melanin skin images. Erythema segmentation is converted into a classification problem by classifying erythema pixels from the non-erythema skin pixels according to the colour intensities in the haemoglobin and melanin skin images. In this section, an SVM is used to perform the classification.

The reason of using the SVM classifier is due to its promising empirical performances [86–88]. This attractive feature results from the usage of the kernel trick, through which data

points are projected into a high dimensional space to generalise the decision rule. Thus a nonlinear classification problem is resolved in a linear way.

SVMs are originally proposed to perform linear classification [116]. In a feature space, a hyperplane is identified to separate data points and has the largest margin. The margin of a hyperplane is the distance of the hyperplane to the closest data points from each side of the space separated by the hyperplane. Later, to resolve non-linear classification problems, the kernel trick is used by defining a kernel function [117]. The kernel function of the inner product of the two vectors in the original low dimensional space is equal to the inner product of any two mapped vectors in the higher dimensional space. By using the kernel trick, the linear hyperplane identified in the higher dimensional space is a non-linear decision boundary in the original low dimensional space. Widely used kernels include polynomials, Radial Basis Functions (RBF) and hyperbolic tangents.

In this work, since there is a sharp disparity between erythema and non-erythema skin in melanin and haemoglobin skin images as shown in Figure 3.5, the feature space is defined using colour components in the two kinds of skin images. For the ith feature vector $\mathbf{v}_i$, we have:

$$\mathbf{v}_i = [(\tilde{I}_i^m)^t, (\tilde{I}_i^h)^t]^t \tag{3.30}$$

where $\tilde{I}_i^m$ is the RGB colour intensity of the melanin component, and $\tilde{I}_i^h$ is the RGB colour intensity of the haemoglobin component.

An illustration of the SVM in the application of erythema classification using a linear classification is shown in Figure 3.6, where two hyperplanes are illustrated: one marked in blue and the other marked in green. The hyperplane in blue that has the maximum margin is the one that is looked for.

Since it is hard to guarantee that the distributions of erythema and non-erythema skin pixels are linearly separable, the feature vectors of the training samples are mapped non-linearly into a higher dimensional space, so that a more suitable separating hyperplane can be found and the classification accuracy can be better. The kernel function we choose is an RBF, due to the simplicity and efficiency for both linear and nonlinear data sets. The advantages of the RBF are because of that: (1) the RBF has only one parameter, which is

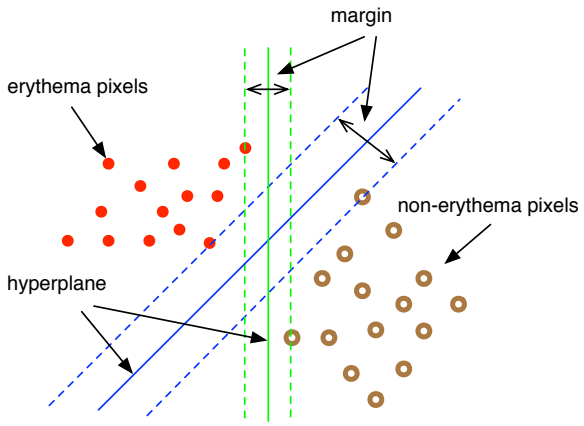


Figure 3.6: Support vector machine classification of erythema pixels from non-erythema pixels in the example of linear separable classification situation.

less than two parameters required in the polynomials and the hyperbolic tangents; (2) the range of the RBF is $[0, 1]$, which makes the computation easier, since the infinity would be achieved by the polynomials and the hyperbolic tangents. A brief introduction of the SVM classifier used in this work is present below.

Given a training set $\mathcal{T}$:

$$\mathcal{T} = \{(\mathbf{v}_1, \alpha_1), ...(\mathbf{v}_i, \alpha_i), ...(\mathbf{v}_n, \alpha_n)\}, \quad \mathbf{v}_i \in \Re^6, \alpha_i \in \{-1, 1\} \tag{3.31}$$

where if $\alpha_i = -1$, $\mathbf{v}_i$ is a feature vector of an erythema pixel, and if $\alpha_i = 1$, $\mathbf{v}_i$ is a feature vector of a non-erythema skin pixel. n is the number of training samples. We have a hyperplane $h(\mathbf{v})$:

$$h(\mathbf{v}) = \mathbf{w}^t\mathbf{v} + b = 0 \tag{3.32}$$

where b is the intercept term and $\mathbf{w}$ is a hyperplane normal vector. $\mathbf{w}$ is perpendicular to the hyperplane and determines the direction of the hyperplane. Finding the separating hyperplane is equivalent to identifying the parameters of the hyperplane: b and $\mathbf{w}$. The margin of the hyperplane to the two classes is scaled as $\frac{2}{\|\mathbf{w}\|}$, so that $h(\mathbf{v}) = 1$ at the nearest point in one classes and $h(\mathbf{v}) = -1$ at the nearest point in the other class. Maximising the margin can be converted to a minimisation problem that is given by:

$$\begin{aligned} \text{minimise} \quad & \frac{1}{2}\mathbf{w}^t\mathbf{w} + c\Sigma_{i=1}^{n}\xi_i \\ \text{subject to} \quad & \alpha_i(\mathbf{w}^t\mathbf{v}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, ..., n \\ & \xi_i \geq 0, \quad i = 1, 2, ..., n \end{aligned} \tag{3.33}$$

where ξ_i is a slack variable and c is an error tolerance constant. They are introduced for the situation where training samples are not completely separable. The slack variable ξ_i allows some data points to be misclassified by the hyperplane. The error tolerance c is positive and controls tolerance of the slack variables. When the two classes are separable, $\xi_i = 0$ $(i = 1, 2, ..., n)$.

Moreover, to resolve a non-linear classification problem, the RBF is defined as:

$$\begin{aligned} K(\mathbf{v}_i, \mathbf{v}_j) &= \phi(\mathbf{v}_i)^t\phi(\mathbf{v}_j) \\ &= \exp(-\frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{2\sigma^2}), \gamma \geq 0 \end{aligned} \tag{3.34}$$

where $\phi(\cdot)$ is the mapping function and σ is a parameter of the RBF kernel, which decides smoothness of the hyperplane.

By projecting a feature sample $\mathbf{v}_i$ to a high dimensional space using the mapping function $\phi(\cdot)$, the objective function of the SVM is rewritten as:

$$\begin{aligned} \text{minimise} \quad & \frac{1}{2}\mathbf{w}^t\mathbf{w} + c\sum_{i=1}^{n}\xi_i \\ \text{subject to} \quad & \alpha_i(\mathbf{w}^t\phi(\mathbf{v}_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, ..., n \\ & \xi_i \geq 0, \quad i = 1, 2, ..., n \end{aligned} \tag{3.35}$$

The minimisation problem given by Eq. (3.35) is a quadratic cost function with linear inequality constraints. This problem can be resolved by maximising the associated dual representation, which is given below:

$$\begin{aligned} \text{maximise} \quad & \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \alpha_i \alpha_j \phi(\mathbf{v}_i)^t \phi(\mathbf{v}_j) \\ \text{subject to} \quad & 0 \leq \lambda_i \leq c, \quad i = 1, 2, ..., n \\ & \sum_{i=1}^{n} \lambda_i \alpha_i = 0 \end{aligned} \tag{3.36}$$

where λ_i is a Lagrange multiplier. By substituting the Eq. (3.34) into the cost function of Eq. (3.36), the mapping function $\phi(\cdot)$ is replaced with the kernel $K(\cdot, \cdot)$. The cost function of Eq. (3.36) is re-expressed as:

$$\text{maximise} \quad \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \alpha_i \alpha_j K(\mathbf{v}_i, \mathbf{v}_j) \tag{3.37}$$

After the maximisation problem is resolved, test samples are classified by the hyperplane in the testing stage. Since we have the hyperplane normal vector $\mathbf{w} = \sum_{i=1}^{n} \lambda_i \alpha_i \phi(\mathbf{v}_i)$, the hyperplane function $h(\mathbf{v})$ is redefined using the kernel $K(\cdot, \cdot)$:

$$\begin{aligned} h(\mathbf{v}) =& \mathbf{w}^t \phi(\mathbf{v}) + b \\ =& \sum_{i=1}^{n} \lambda_i \alpha_i \phi(\mathbf{v}_i)^t \phi(\mathbf{v}) + b \\ =& \sum_{i=1}^{n} \lambda_i \alpha_i K(\mathbf{v}_i, \mathbf{v}) \end{aligned} \tag{3.38}$$

For a testing sample $\mathbf{v}$, the sign of the function $h(\mathbf{v})$ determines the class that the sample $\mathbf{v}$ belongs to.

3.5 Experimental results: erythema segmentation

In this section, the haemoglobin and melanin components estimated by ICA are validated at first by comparing with the extinction coefficients of melanin and haemoglobin. This is followed by a verification of the SVM classification performance. The superiority of our algorithm is shown by comparing with a nearest neighbour classification.

3.5.1 Validation of the haemoglobin and melanin components

Independent Component Analysis (ICA) is ill-posed. Without accurate knowledge of the mixing process of the melanin component and the haemoglobin component, the skin colour decomposition is indeterminate. In other words, it is hard to guarantee that the signals extracted by ICA are the exact haemoglobin and melanin components. For skin regions with different skin colours, the independent components derived by ICA are different. In some cases, especially when other skin traits, such as moles, dominate the skin region, the estimated independent components do not correspond to the melanin and haemoglobin components. It is possible that the melanin or haemoglobin components are discarded in the ICA process, because they do not occupy the majority of the region.

To solve the problem, we use ICA to decompose a skin image only containing erythema and normal skin in order to estimate the haemoglobin component $\mathbf{c}^{\tilde{h}}$ and the melanin component $\mathbf{c}^{\tilde{m}}$. The estimated melanin component $\mathbf{c}^{\tilde{m}}$ and haemoglobin component $\mathbf{c}^{\tilde{h}}$ are then used to decompose other erythema skin images, including those containing other skin traits. In this research, the estimated haemoglobin component $\mathbf{c}^{\tilde{h}}$ and melanin component $\mathbf{c}^{\tilde{m}}$ are specified as:

$$\tilde{C} = [\mathbf{c}^{\tilde{m}}, \mathbf{c}^{\tilde{h}}] = \begin{bmatrix} c^{\tilde{m}}(1) & c^{\tilde{h}}(1) \\ c^{\tilde{m}}(2) & c^{\tilde{h}}(2) \\ c^{\tilde{m}}(3) & c^{\tilde{h}}(3) \end{bmatrix} = \begin{bmatrix} 0.0246 & 0.0193 \\ 0.0316 & 0.0755 \\ 0.0394 & 0.0666 \end{bmatrix} \tag{3.39}$$

where $\mathbf{c}^{\tilde{m}}$ is the estimated melanin component, $\mathbf{c}^{\tilde{m}} = [c^{\tilde{m}}(1), c^{\tilde{m}}(2), c^{\tilde{m}}(3)]^t$, and $\mathbf{c}^{\tilde{h}}$ is the estimated haemoglobin component, $\mathbf{c}^{\tilde{h}} = [c^{\tilde{h}}(1), c^{\tilde{h}}(2), c^{\tilde{h}}(3)]^t$.

Since in this work, we consider the melanin and haemoglobin components as the intrinsic light absorption properties of melanin and haemoglobin pigments, validation of the estimated melanin and haemoglobin components, $\mathbf{c}^{\tilde{m}}$ and $\mathbf{c}^{\tilde{h}}$, is conducted by examining the corresponding absorption spectra of the molar extinction coefficients [1]. In Figure 3.7, the changes of extinction coefficients of melanin and haemoglobin versus the wavelength values are shown separately. It is observed that the extinction coefficients of melanin and haemoglobin have limited value ranges with respect to the wavelengths of

[1] measurements of the degree of light absorption of a substance associated with a wavelength

visible light. Additionally, it is pointed out in [118] that values of the melanin and haemoglobin components are proportional to their respective extinction coefficients. In the rest of this section, we validate the estimated melanin and haemoglobin components by examining their respective absorption spectra in the wavelength ranges of red, blue and green separately.

Validation of the estimated melanin component

In Figure 3.7, the melanin extinction coefficient is illustrated with the extinction coefficients of two kinds of melanin together, the brown-black eumelanin and the red-brown phaeomelanin. The extinction coefficient of melanin monotonically decreases with the wavelength. Specifically for the wavelength ranges (Red=[622-760], Green=[492-577] and Blue=[455-470]), the extinction coefficient of melanin decreases from the blue wavelength range to the red wavelength range. Xu et al. [118] concluded that:

$$\epsilon_R^m \in [2,4], \quad \epsilon_G^m \in [5,7], \quad \epsilon_B^m \in [8,10] \tag{3.40}$$

where ϵ_i^m denotes the extinction coefficient of melanin at specified wavelength range i; R, G, and B represent spectral regions of red, green and blue respectively.

Furthermore, Xu et al. derived a rule that:

$$V_R^m \in [0.1,0.5], \quad V_G^m \in [0.3,0.7], \quad V_B^m \in [0.7,0.1] \tag{3.41}$$

where V_R^m, V_G^m and V_B^m are normalised melanin components with respect to the red, green and blue wavelengths respectively.

Normalising $\tilde{\mathbf{c}^m}$, we have:

$$\overline{\tilde{\mathbf{c}^m}} = [0.4379, 0.5625, 0.7013]^t \tag{3.42}$$

where $\overline{\tilde{\mathbf{c}^m}}$ is the normalised $\tilde{\mathbf{c}^m}$. The elements in $\overline{\tilde{\mathbf{c}^m}}$, $\overline{\tilde{c^m}}(1)$, $\overline{\tilde{c^m}}(2)$ and $\overline{\tilde{c^m}}(3)$, correspond to the wavelength of red, green and blue respectively and fit the requirement in Eq. (3.41).

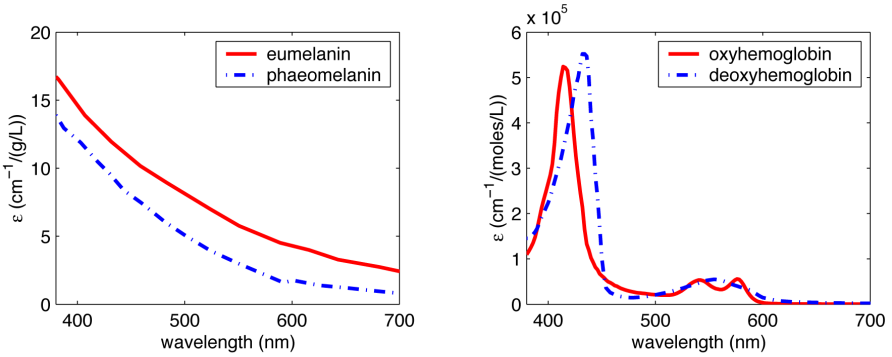


Figure 3.7: Spectral extinction coefficient curves for melanin and haemoglobin pigments in the skin [119].

Validation of the estimated haemoglobin component

With regard to the haemoglobin component, no obvious rules can be drawn other than that ϵ_R^h is not the biggest one [118]. In Figure 3.7, haemoglobin is presented using oxygenated haemoglobin, oxyhemoglobin, and deoxygenated haemoglobin, deoxyhemoglobin, together. It shows that the extinction coefficient of the haemoglobin component has a higher value in the spectral ranges of blue and green than the red wavelength range.

For the $\tilde{\mathbf{c}^h}$ used in the research, we have:

$$\tilde{c^h}(1) < \tilde{c^h}(2), \quad \tilde{c^h}(1) < \tilde{c^h}(3) \tag{3.43}$$

where $\tilde{c^h}(1)$, $\tilde{c^h}(2)$ and $\tilde{c^h}(3)$ are corresponding to the wavelengths of red, green and blue respectively. They fit the experimental observation in Figure 3.7.

By using the validated haemoglobin and melanin components, a skin colour is decomposed and expressed with the melanin and haemoglobin quantities. Additionally, a fixed parameter $\mathbf{d}$ is used in the skin colour decomposition. It is also derived the skin image containing erythema and normal skin only. We have $\mathbf{d} = [1.1601, 2.8347]^t$ in this research.

3.5.2 Evaluation of the erythema segmentation

To test the performance of the SVM based erythema segmentation method, we selected 100 images from the database, including various skin colours and erythema severity. These images are collected from Skin & Cancer Foundation Victoria, and are captured indoors with a controlled illumination. Erythema lesions in the images are outlined by a dermatologist as ground truth. We divided the database into a 40-image training set, where 150000 erythema pixels and 170000 non-erythema pixels are extracted, and a testing set consisting of the remaining 60 images.

The LIBSVM library is used to build the SVM classifier [120]. We set the error tolerance constant $c = 1000$ and the parameter of the RBF kernel $\sigma^2 = 5$ in this research. The parameters σ and c is determined through a "grid-search" using a cross-validation [120]. Candidates of σ and c increase exponentially and a 3-fold cross-validation is applied. This process divides the training set into three subsets, where one of the subsets is chosen in turn for testing and the rest are used for training. The pair σ and c is chosen, when the best cross-validation accuracy is achieved, which is the percentage of correctly classified data in the training set.

To evaluate the performance of the SVM classification, sensitivity and specificity are used in this work. The sensitivity is a measurement of true positive rate, and the specificity is a measurement of true negative rate:

$$\begin{aligned} \text{sensitivity} &= \frac{TP}{TP + FN} \\ \text{specificity} &= \frac{TN}{TN + FP} \end{aligned} \tag{3.44}$$

where TP, FN,TN and FP are the numbers of true positives, false negatives, true negatives and false positives respectively. In the erythema segmentation research, the sensitivity means the percentage of pixels that do belong to erythema in the total classified erythema pixels, and the specificity is the percentage of pixels that are non-erythema skin in the total classified non-erythema skin pixels.

The performance of the SVM classification based segmentation method is compared

with the nearest neighbour classification method proposed in [11]. [11] segments vitiligo, a skin depigmented disorder due to the lack of melanin, from normal skin by using ICA to decompose skin into the melanin and haemoglobin components as well. However, in [11], a nearest neighbour classification method is employed to perform the segmentation. A pixel is labelled as vitiligo, if the Euclidean distance of its melanin quantity to vitiliginous skin samples is closer than to healthy skin samples.

The method used in [11] is amended by calculating the Euclidean distance of colour vectors in haemoglobin skin images to separate erythema from healthy skin, since erythema is accompanied by an increase of haemoglobin pigment. Two pigment vectors $\mathbf{a}_e$ and $\mathbf{a}_s$ are introduced, which are the means of erythema colour intensities and non-erythema skin colour intensities respectively in the haemoglobin skin images from the training set. The Euclidean distances between colour vectors in the haemoglobin skin images are calculated as below:

$$\begin{aligned} d_e &= \|\mathbf{h}_{x,y} - \mathbf{a}_e\| \\ d_s &= \|\mathbf{h}_{x,y} - \mathbf{a}_s\| \end{aligned} \tag{3.45}$$

where d_e is the Euclidean distance between a colour vector $\mathbf{h}_{x,y}$ and the mean of erythema colour intensities $\mathbf{a}_e$, d_s is the Euclidean distance between a colour vector $\mathbf{h}_{x,y}$ and the mean of non-erythema skin colour intensities $\mathbf{a}_s$, and $\|\cdot\|$ is the Euclidean norm. If $d_e \leq d_s$, the pixel is classified as erythema. Otherwise, the pixel is identified as a non-erythema pixel.

Table 3.1 shows the sensitivity and specificity of our proposed algorithm together with the nearest neighbour classification method used in [11]. Even though the nearest neighbour classification method has a higher specificity value, the SVM-based method presented in this chapter shows a higher sensitivity value. In the dermatology area, it is required to have a higher sensitivity than other metrics. Since the sensitivity measures the percentage of true positives of classified erythema pixels, while the specificity focuses on the measurement of the true positives of classified non-erythema pixels, a higher sensitivity indicates a higher classification performance for erythema, including

mild and severe erythema. In this case, a lower sensitivity means the classification rate for mild erythema is lower. A high sensitivity evaluation ensures that the erythema area can be evaluated accurately and the progress of the treatment can be presented clearly. Figure 3.8 shows experimental results of a subset of images. It is observed that the SVM-based method is able to segment more erythema pixels with minor severity degree than the nearest neighbour classification method.

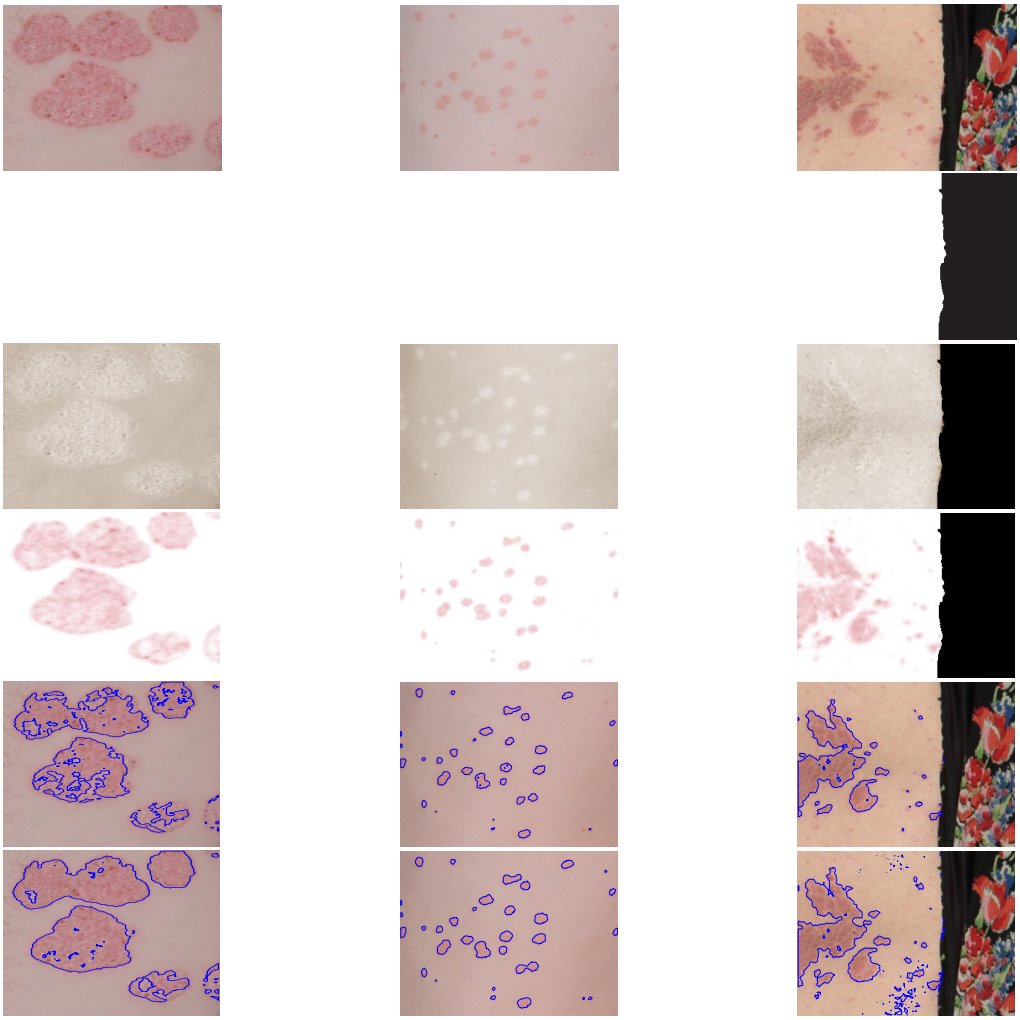
The lower specificity or our algorithm is mainly due to shadows on the skin or gradually darkening skin. Since the haemoglobin Cb and Cr quantities in shadow and in darkening skin are close to erythema, shadows and darkened skin would be misclassified as erythema. Some moles are misclassified as erythema with both our algorithm and the method in [11], which are parameterless methods based on K-means clustering. This misclassification is because red pigments sometimes appear around moles (as shown in the second column of Figure 3.8).

Methods	Sensitivity	Specificity
Methods in [11]	89.41%	85.34%
Our methods with the SVM	95.32%	75.01%

Table 3.1: Performance of algorithms.

3.6 Summary

In this chapter, a novel scheme to segment erythema from psoriasis skin images is proposed. After skin pixels are detected in the psoriasis skin images, skin colour is represented as a linear combination of the melanin component and the haemoglobin component, which are derived using ICA. A trained SVM is employed to separate erythema based on the distribution of colour intensities on the melanin and haemoglobin skin images. The sensitivity of this algorithm suggests that this proposed method is applicable in clinical treatments.



From top to bottom: original image, skin segmentation result, melanin component image, haemoglobin component image, segmentation result with method in [11] and our segmentation result.

Figure 3.8: Examples of the erythema segmentation result.

Chapter 4
Scaling Segmentation

4.1 Introduction

The appearance of scaling varies with different psoriatic lesions. It typically appears as white or creamy coloured scales on regions of red and inflamed skin (erythema) but can also appear in isolation without the accompanying erythema. When psoriasis appears without the accompanying erythema it appears as discernibly white or creamy flakes on normal skin. Scaling can present as small spots or as patches scattered within erythema. Figure 4.1 shows some examples of the variation in the appearance of scaling. This variation makes it difficult to identify scaling boundaries through more conventional boundary detection algorithms and as a consequence we use a pixel based classification and labelling approach.

Moreover, the colour of scaling may be very similar to that of normal skin, especially if the skin is fair, making it difficult to differentiate between scaling and normal skin using colour alone. However, the rough textured surface of scaling is markedly different from normal skin. The algorithm presented in this chapter uses a feature space derived from both colour and texture to identify scaling pixels.

The algorithm for scaling detection is a pipeline that is essentially a pixel labelling algorithm that identifies scaling in 2D digital skin images without the need for locating psoriasis first. It is composed of two main stages: (1) a feature extraction stage, and (2) a scaling segmentation stage. The two stages are described as follows (See Figure 4.2) :

Step 1: The algorithm first analyses skin colour and skin texture using an appropriately chosen colour space and a bank of Gabor filters to create a feature space for the image

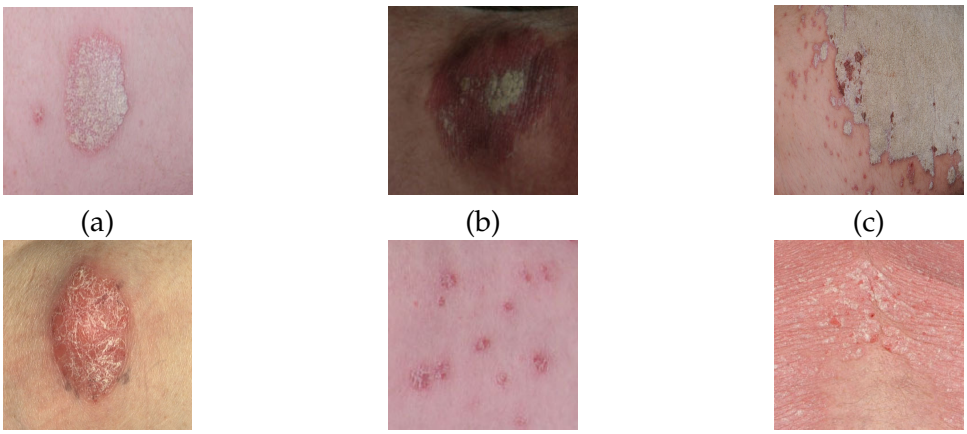


Figure 4.1: Examples of scaling in psoriasis lesions. (a) Scattered scaling in plaque psoriasis; (b) Patched scaling in plaque psoriasis; (c) Extensively covered scaling in plaque psoriasis; (d) Scaling in guttate psoriasis; (e) Scaling in pustular psoriasis; and (f) Scaling in erythrodermic psoriasis.

(see Section 4.2).

Step 2: The algorithm next removes erythema pixels from consideration and resamples the image to collect training samples for the classification process. The segmentation is achieved by using a combination of a Markov Random Field (MRF) and the hyperplane derived from a Support Vector Machine (SVM) (see Section 4.3).

This chapter begins with a derivation of a feature space for the analysis of scaling. In Section 4.2 we design a multi-window filter for detecting colour contrast between scaling and erythema. We also design a bank of Gabor filters to describe the textures that correlate strongly with scaling.

Secondly, in Section 4.3 the scaling segmentation method is described. A classifier that combines the hyperplane from an SVM with an MRF is developed to first classify pixels and then smooth the result by taking the spatial context into account. Section 4.3 also describes how the training sets are collected directly from the image being analysed through a soft-constrained k-means.

Thirdly, in Section 4.4 the segmentation results and experimental validation of the algorithms are presented. The performance of training set collection algorithm is examined, followed by a comparison between the proposed classifier with an SVM alone and an MRF alone. Parameters used in the classifier are examined as well.

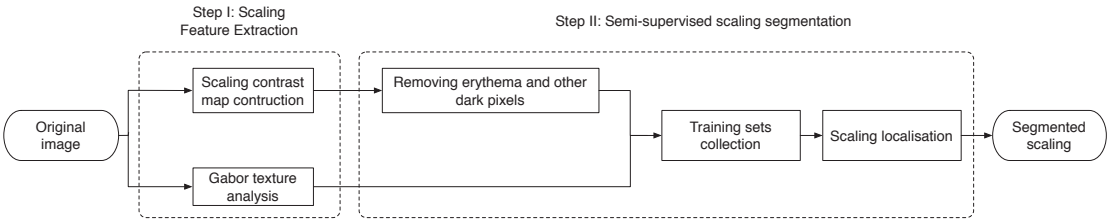


Figure 4.2: A flow chart of the algorithm for automatic segmentation of scaling in 2D psoriasis skin images.

4.2 Feature space for scaling detection

4.2.1 Scaling Contrast Map Construction

In this section, a scaling contrast map is developed to enhance the contrast between scaling and erythema. The map aims to enhance the contrast of scaling especially in situations where scaling is scattered in erythema and is hard to discern visually. The L*a*b* colour space is used to develop a pair of multi-window filters that increases the contrast between scaling and erythema. To speed up the calculation, an integral image is applied in the multi-window filtering process.

The L*a*b* colour space

The L*a*b* colour space is a non-linear transformation from the CIE XYZ space. The CIE XYZ space was created by the Commission Internationale de l'Eclairage (CIE), an International commission that develops standards for light and colour. This colour space is derived from a series of psychophysical experiments to match human colour perception. A conversion from the RGB colour space to the CIE XYZ with the reference colour of white D65 is expressed below [121]:

$$\begin{aligned} \mathbf{X} &= 0.412\mathbf{R} + 0.358\mathbf{G} + 0.18\mathbf{B} \\ \mathbf{Y} &= 0.213\mathbf{R} + 0.715\mathbf{G} + 0.072\mathbf{B} \\ \mathbf{Z} &= 0.019\mathbf{R} + 0.119\mathbf{G} + 0.95\mathbf{B} \end{aligned} \tag{4.1}$$

In the L*a*b* colour space, the colour distance is linearised for the perception of colour difference. In other words, the perceptual difference of two colours is able to be described by their distance in the colour space. The transformation from the CIE XYZ colour space to the L*a*b* colour space is expressed as [122]:

$$\begin{aligned} \mathbf{L}^* &= 116 f(\frac{\mathbf{Y}}{Y_n}) - 16 \\ \mathbf{a}^* &= 500[f(\frac{\mathbf{X}}{X_n}) - f(\frac{\mathbf{Y}}{Y_n})] \\ \mathbf{b}^* &= 200[f(\frac{\mathbf{Y}}{Y_n}) - f(\frac{\mathbf{Z}}{Z_n})] \end{aligned} \tag{4.2}$$

where

$$f(t) = \begin{cases} t^{\frac{1}{3}} & t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & otherwise \end{cases} \tag{4.3}$$

The X_n, Y_n and Z_n are the colour values of the reference white in the CIE XYZ colour space. The L*a*b* colour space is relative to the reference white colour and is thus device-independent.

The $\mathbf{L}^*$ dimension specifies lightness where an $\mathbf{L}^*$ value of 0 is black and an $\mathbf{L}^*$ value of 100 is a diffuse white. The $\mathbf{a}^*$ dimension is the red-green dimension, where a positive value of $\mathbf{a}^*$ is red and a negative value is green, and the $\mathbf{b}^*$ dimension is the blue-yellow dimension, where a positive value of $\mathbf{b}^*$ is blue and a negative value is yellow.

Using multi-window filtering to construct a scaling contrast map

The colour of scaling correlates well with higher values of $\mathbf{L}^*$ and erythema with positive values of $\mathbf{a}^*$. Shadows result in smaller $\mathbf{L}^*$ values but do not necessarily affect the other dimensions. With this in mind a **scaling contrast map** S is defined as the result of multi-window filtering in the $\mathbf{L}^*$ component and the $\mathbf{a}^*$ component:

$$S_{x,y} = J(L^*_{x,y}) + J(\text{inv}(a^*_{x,y})) \tag{4.4}$$

where $S_{x,y}$ is the value of scaling contrast map S at the image coordinate (x, y), $J(\cdot)$ is a multi-window filter that detects contrast in the specific colour component, and $\text{inv}(a^*_{x,y})$ inverts the image in the a^* dimension, which is defined by:

$$\text{inv}(a^*_{x,y}) = \max_{i,j}(a^*_{i,j}) - a^*_{x,y} \tag{4.5}$$

where (i, j) runs through all the coordinates in the image. Through the inverting operation, the colour intensity values of erythema become low and the scaling become high. It possesses a similar property with the colour intensity values in the L^* component. The combination of contrast detection in the $\mathbf{L}^*$ component and the $\mathbf{a}^*$ component enhances the contrast between scaling and the surrounding erythema.

The multi-window filter $J(\cdot)$ defined by:

$$J(X_{x,y}) = \sum_{s=1}^{3}(X_{x,y} - \frac{1}{N}\sum_{\substack{x-w(s)\leq m\leq x+w(s)\\ y-w(s)\leq n\leq y+w(s)}} X_{m,n}) \tag{4.6}$$

where $X_{x,y}$ is the colour intensity of the current pixel with coordinate (x, y) and s is the scale. The subtracted term is the average intensity value of the surrounding area with pixel number N. The size of the surrounding area is determined by scale s:

$$w(s) = \frac{d}{2^s} \tag{4.7}$$

where d is the larger value between the image width and the image height, and the scale s is set to be $s \in \{1, 2, 3\}$ [123]. These scale values cover the contrast analysis from local information to global information.

Using an integral image to speed up the multi-window filtering

The multi-window filtering requires to calculate the intensity value of the surrounding area for each pixel in the image. It is time consuming to do this calculation directly. An integral image is used to simplify this process. In the integral image, a pixel value $ii_{x,y}$ at location (x, y) contains the sum of pixel values above and to the left [124].

$$ii_{x,y} = \sum_{x' \leq x,\ y' \leq y} X(x', y') \tag{4.8}$$

The integral image is derived by integrating pixel values in y direction followed by integrating in x direction.

$$\begin{aligned} s_{x,y} &= s_{x,y-1} + X_{x,y} \\ ii_{x,y} &= ii_{x-1,y} + s_{x,y} \end{aligned} \tag{4.9}$$

Given the multi-window filter $J(X_{x,y})$ (see Eq. (4.6)), the summation of pixel values in the surrounding area of $X_{x,y}$ is simply achieved through the calculation of pixel values at four conners of the surrounding area in its integral image:

$$\sum_{\substack{x-w(s) \leq m \leq x+w(s) \\ y-w(s) \leq n \leq y+w(s)}} X_{m,n} = ii_{x-w(s),y-w(s)} + ii_{x+w(s),y+w(s)} - ii_{x-w(s),y+w(s)} - ii_{x+w(s),y-w(s)} \tag{4.10}$$

Using the integral image to compute the multi-window filtering is faster than the traditionally used Fast Fourier Transformation (FFT). According to Eq. (4.9), the computation complexity of the integral image is $O(mn)$, where m is the width of the image and n is the height of the image. It is lower than the computation complexity of FFT having the value of $O(mnln(mn))$ [125]. Figure 4.3 illustrates using the integral image to calculate the summation of pixel values in the grey block with four corners: A, B, C, and D.

The contrast filter $J(\cdot)$ compares the intensity of the current pixel with its surroundings at the different scales, so that the scaling becomes conspicuous in the image. In normal lighting conditions, if (x, y) is a scaling pixel, then $J(L^*_{x,y})$ is positive. In the presence of shadows $J(L^*_{x,y})$ is still positive, providing there is sufficient contrast with the surrounding pixels. In the case where shadow lowers the colour intensity of the $\mathbf{L}^*$ component to give a negative value of $J(L^*_{x,y})$, the inverted $\mathbf{a}^*$ component compensates for the shadows, since the $\mathbf{a}^*$ component is less susceptible to the uneven illuminance. The $J(\cdot)$ filter does this by inverting the $\mathbf{a}^*$ values in the surrounding pixels which again ac-

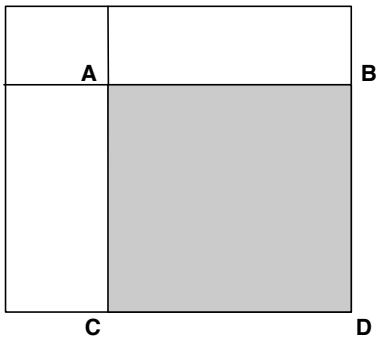
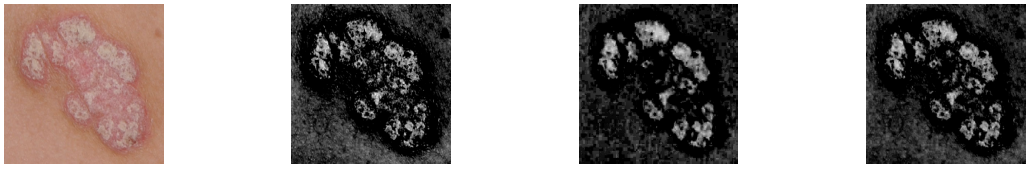


Figure 4.3: An integral image: the summation of pixel values in the grey block of the corresponding original image has the value of $ii_A + ii_D - ii_C - ii_B$.



(a) (b) (c) (d)

Figure 4.4: Scaling contrast map construction. (a) Original image; (b) Contrast map derived from L^*; (c) Contrast map derived from a^*; (d) Scaling contrast map.

centuates the positive difference in contrast between the current centred pixel and the surroundings. In this way the scaling contrast map S is robust to the change of illuminance. An example of the construction is shown in Figure 4.4.

4.2.2 Texture analysis with Gabor filters

The scaling contrast map S behaves well when segmenting scaling from erythema but is not sufficient for segmenting scaling from normal skin, especially when the colour difference between the two is small. However, scaling presents as a rough textured surface in 2D images that distinguishes it from the more smoothly textured normal skin. The rough texture of scaling combined with the scaling contrast map is able to provide a good combination of features for segmenting scaling.

Though there is no definite rule to describe the texture of scaling, the appearance of scaling is rough and this roughness is visually apparent in the psoriasis skin images through changes in colour intensity. In this work, Gabor filters are used to analyse the

roughness of the scaling texture. A bank of Gabor filters are designed to differentiate different kinds of scaling from normal skin. The filtering results are fused into a grey scale Gabor texture image, in which rough scaling has a higher intensity value than the normal skin.

Gabor filters and Gabor energy

Gabor filters have a long history in the analysis of image textures. In [126], a bank of Gabor filters are designed to classify different patterns of textures; in [127], a Gabor filter is optimally tuned for a texture segmentation. In these papers, the textures are periodic patterns with changes of colour intensity. More generally, when changes of colour intensity are not periodic and the rule for the changes is hard to determine, Gabor filters are still able to detect the texture features. The performance of the Gabor filter for non-periodic patterns is tested in apple quality inspection [128] and face recognition [129]. Keeping this in mind, we use Gabor filters in this work to differentiate non-periodic scaling patterns from normal skin patterns.

The Gabor filter was originally proposed by Gabor in 1946 [130] and is later extended to two dimensions by Daugman [131]. A 2D Gabor filter is defined as:

$$\begin{aligned} g(x,\, y; \gamma,\, \sigma,\, \lambda, \psi) &= \exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}) \exp(i(2\pi \frac{x'}{\lambda} + \psi)) \\ x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned} \tag{4.11}$$

where the Gaussian distribution function $\exp(-(x'^2 + \gamma^2 y'^2)/2\sigma^2)$ with standard deviation σ and spatial aspect ratio γ is called the envelop, the complex sinusoidal $\exp(i(2\pi x'/\lambda + \psi))$ with spatial frequency $1/\lambda$ and phase shift ψ is called the carrier, and θ is the rotation angle. An example of a 2D Gabor filter is shown in Figure 4.5.

The *response* of the Gabor filter $r_{x,y}$ is defined via convolution of the filter $g(x, y)$ with the grey scale image G:

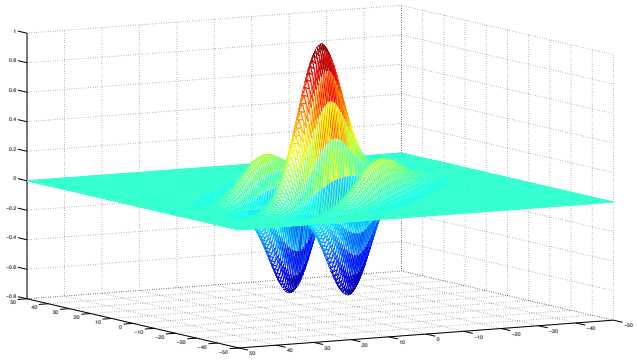


Figure 4.5: A 2D Gabor filter with $\theta = \frac{\pi}{4}$, $\gamma = 0.5$, $\psi = 0$ and $h = 1$.

$$\begin{aligned} r_{x,y} &= \iint_{\Omega} G(\mu,\eta) g(x-\mu, y-\eta)\, d\mu\, d\eta \\ &= \mathrm{Re}(r_{x,y}) + i\mathrm{Img}(r_{x,y}) \end{aligned} \tag{4.12}$$

where Ω is the set of image points. The response has both real and complex parts: $\mathrm{Re}(r_{x,y})$ and $\mathrm{Img}(r_{x,y})$.

For a fast computation the Fourier transformation is applied to calculate the convolution. Let $\mathcal{F}(\cdot)$ denote an application of the Fourier transformation and $\mathcal{F}^{-1}(\cdot)$ be an inverse Fourier transformation. According to the convolution theorem, we have:

$$\mathcal{F}(r) = \mathcal{F}(I) \cdot \mathcal{F}(g) \tag{4.13}$$

And the Gabor response $r_{x,y}$ is rewritten using the inverse Fourier transformation:

$$r_{x,y} = \mathcal{F}^{-1}(\mathcal{F}(r))_{x,y} \tag{4.14}$$

Through the Fourier transformation, the computational complexity of the convolution is decreased to a pixel-wise multiplication. The Fast Fourier transformation is used in this research to further speed up calculation [132].

Generally, the Gabor energy is used to represent the response of a Gabor filter. The Gabor energy $E_{x,y}$ is defined as the magnitude of the Gabor filter response [126]:

$$E_{x,y}^2 = \mathrm{Re}(r_{x,y})^2 + \mathrm{Img}(r_{x,y})^2 \tag{4.15}$$

We use the square of the Gabor energy because it is better in accentuating the differences between scaling and normal skin than the more commonly used Gabor energy. The response is *highest* when the image intensity frequency is close to the Gabor filter. For smooth normal skin the image intensity is relatively homogeneous and is not sensitive to Gabor filters. For rougher scaly skin, the change in intensity is relatively higher.

Defining a bank of Gabor filters

The variations in the textures of scaling and normal skin in different lesions and in different people make the choice of one single Gabor filter unlikely. The algorithm uses a bank of 24 Gabor filters designed to respond well in a variety of skin and scaling texture conditions. Parameters for the Gabor filters are shown in Table 4.1. The frequencies $1/\lambda$ and rotations θ are chosen in the purpose of covering the spatial-frequency domain appropriately [133]. The choice of other parameters are described below:

The choice of the standard deviation σ of the Gaussian envelope is related with the half-response spatial-frequency bandwidth h and the spatial frequency $1/\lambda$. The relationship between the ratio $\frac{\sigma}{\lambda}$ and the spatial frequency bandwidth is expressed as [134]:

$$\frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^h+1}{2^h-1} \tag{4.16}$$

The spatial frequency bandwidth corresponds to the number of parallel excitatory and inhibitory stripe zones in a receptive field. Though the spatial frequency bandwidth varies in a range that is above and below one octave according to neurophysiological researches, it is pointed out that the information provided by the narrowly tuned simple cells with around one octave half-response spatial frequency bandwidth will be sensed to process [135, 136]. Setting the bandwidth $h = 1$, we have $\sigma = 0.56\lambda$.

The spatial aspect ratio γ describes how elliptical the Gaussian distribution is, and therefore the degree of ellipticity of the receptive field. It is pointed out in [137] that the spatial aspect ratio in the range of $(0.23, 0.92)$. In this thesis, we set the spatial aspect ratio

Table 4.1: Parameters defining the bank of Gabor filters used for scaling texture analysis.

Spatial frequencies	$1/\lambda$	23, 31, 47 cycles per image
Rotation angles	θ	$0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}, \pi$
Phase shift	ψ	0
Spatial aspect ratio	γ	0.5

as a constant $\gamma = 0.5$.

The phase shift $\psi = 0$ specifies the real part of the Gabor function to be symmetric and the imaginary part of the Gabor function to be antisymmetric. The Gabor energy $E_{x,y}$ is an asymmetric mixture.

Obtaining a Gabor texture image

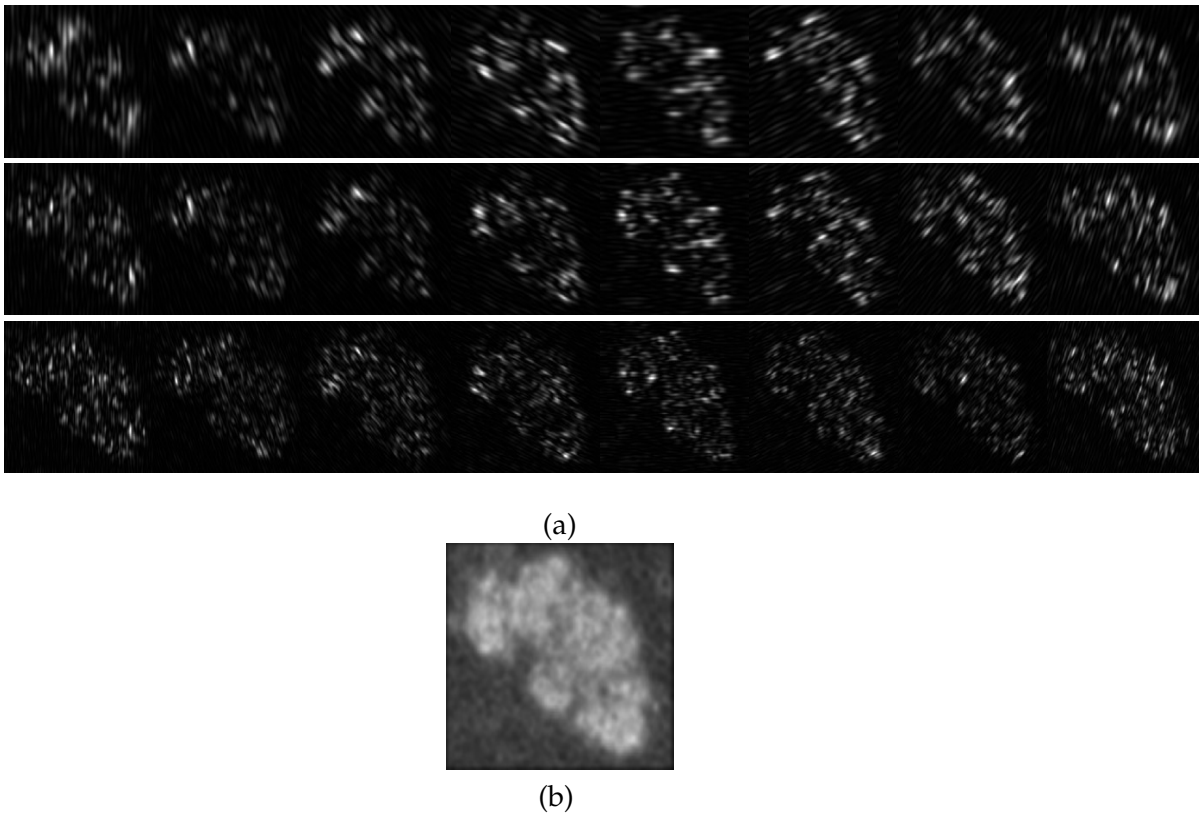
A bank of Gabor filters with parameters given in Table 4.1 are applied to a grey scale psoriasis skin image; accordingly, the corresponding Gabor energy images are obtained. A Gabor texture image is constructed by integrating the Gabor energy images with the purpose of covering the responses of the Gabor filters whose spatial frequencies and rotation angles are between the selected ones in Table 4.1. The technique given in [138] is applied. It comes with the following steps:

1. First, the square of a Gabor energy image is filtered using a hyperbolic tangent to narrow the range. The filtering function is experimentally tuned and expressed as:

$$\tanh(0.25E_{x,y}^2) = \frac{1 - e^{-0.5E_{x,y}^2}}{1 + e^{-0.5E_{x,y}^2}} \tag{4.17}$$

   where parameter 0.25 is derived from experiment. This step functions as a thresholding and prohibit the high response of Gabor filters to some extent.
2. A mean filter is then used to smooth the images with a window size equal to wavelength of the current Gabor filter.
3. Finally, the Gabor texture image is obtained by summing the smoothed output over all of the rotation angles and frequencies of the Gabor filters.

The method of [138] is better at differentiating between scaling and normal skin than simply summing the Gabor energy images directly. The technique accentuates the tex-



(a)

(b)

Figure 4.6: Texture examination corresponding to the original image in Figure 4.4. (a) Gabor filtering responses from a bank of Gabor filters (the spatial frequency changes along the row and the rotation angle changes along the column); (b) The final Gabor feature image.

tures whose orientation and frequency are between the orientations and frequencies in the bank of Gabor filters and suppresses the response of the textures that are beyond the frequency.

An example of a Gabor texture image is shown in Figure 4.6. Scaling has a high Gabor filter response while normal skin has a markedly lower Gabor filter response. The summation in the final step preserves the differences between the higher response from scaling and the lower response from normal skin.

4.3 Semi-supervised scaling segmentation

The second stage of the algorithm segments scaling from 2D skin images through a semi-supervised algorithm to ensure the invariance of segmentation to scaling and skin changes from different patients. This part of the algorithm implements a three step process where:

1. **First**: The scaling contrast map is applied to the image and the resulting image is processed to threshold out all dark pixels representing darker pigments in the skin and including erythema, hair, moles and other blemishes;
2. **Second**: A training set for the scaling classifier is extracted from the image where the training set is composed of pixels that are highly likely to be scaling and pixels that are highly likely to be normal skin; and
3. **Third**: The pixels are classified using an SVM defined by the training set and the resulting image smoothed using an MRF.

4.3.1 Removing erythema and other dark pixels

The first step is to threshold out the dark pixels representing erythema, hair, moles and other blemishes in the scaling contrast map S. Scaling and normal skin pixels remain in consideration after the application of the contrast map because they result in a significantly high value of S. We define a binary image M by:

$$M_{x,y} = \begin{cases} 1 & \text{if } S_{x,y} \geq t_s \\ 0 & \text{otherwise} \end{cases} \tag{4.18}$$

where t_s is the threshold value (see Section 4.4.3 for the validation). Pixels labelled with 1 including scaling and normal skin are retained for further analysis. Pixels labelled with 0 denote darker pigments. Besides erythema, hair, moles and other blemishes are removed from further consideration by the thresholding. This step changed the scaling segmentation problem into a binary classification problem, since only scaling and normal skin remain in the image. In the following steps, a classifier is designed to differentiate scaling from normal skin.

4.3.2 Training data collection

The removal of erythema and darker pixels using Eq. (4.18) simplifies the problem of detecting scaling to a binary classification problem: that of distinguishing scaling from normal skin. Since the Gabor filter is a spatial filter, it causes pixels around borders of a psoriatic lesion to have similar responses with scaling pixels. Pixel-based classifiers, such as the SVM, are not applicable. In this situation, we put the image structure into consideration. An MRF that models the whole image as a graph is applied. The likelihood function of the MRF is derived from the distance of a pixel to the hyperplane of an SVM. The parameters defining the placement of the hyperplane in the feature space need to be derived using carefully chosen training data.

There is a great deal of variation in skin colours and psoriasis lesions. A hyperplane using parameters derived from a generic set of training data gathered over a wide range of images is unlikely to yield good classification results. Our algorithm gathers the training data needed to place the SVM hyperplane directly from the image being analysed. Training data is collected by identifying the regions of scaling and normal skin using the position of erythema, which is often found between scaling and normal skin. Collecting training data proceeds with three steps:

1. First, the erythema location is approximately localised based on the scaling contrast map.
2. Second, according to the erythema location, potential samples of scaling and normal skin pixels are identified with a series of morphological operations.
3. Third, a soft-constrained k-means clustering is used to further identify candidate training regions of scaling and normal skin from the potential samples.

An approximate localisation of erythema for the training set identification

The approximate location of erythema is identified by the grey-scale intensity using the scaling contrast map S, where low grey scale intensities indicate red erythema pixels. A rough segmentation of erythema, but one that serves our purposes, can be obtained through a threshold method:

$$X_{x,y} = \begin{cases} 1 & \text{if } S_{x,y} \leq 0.2 \min_{i,j} S_{i,j} \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

where X is a binary image, which indicates location of erythema with value '1' (white colour in the image), and other objects with value '0' (black colour in the image).

The threshold value is determined empirically, based on the fact that darkened normal skin and erythema would show negative values in the scaling contrast map, but the values of darkened normal skin would still be greater than erythema. Through Eq. (4.19), the approximate location of erythema is identified. It is noted that not all the erythema can be identified through this method, but normal skin and scaling are not included in the identified erythema location. The simple thresholding method ensures the high sensitivity of erythema detection. Thus in the following steps, the training sets of scaling and normal skin can be well identified according to the erythema location.

Identifying potential samples of scaling and skin pixels

The next step is to use the approximate localisation of erythema to identify potential samples of scaling pixels and normal skin samples as the training samples. Using the fact that scaling is often surrounded, or partially surrounded, by erythema, we use dilation and erosion operations to create regions of scaling enclosed by boundaries of erythema. Regions within the boundaries thus created are filled using a flood fill operation.

Scaling is located at the intersection of the white colour regions in the binary image M and the regions in the image X that have been bounded and flood-filled. Normal skin occurs at the intersection of M and the regions that have not been flood filled. Note that it is hard to guarantee that pixels designated as scaling are all scaling. This is because very small regions of skin often occur inside regions of psoriasis. The same can be said for the pixels designated as normal skin which may contain scaling, especially when the scaling is not surrounded by erythema.

The algorithm given in Algorithm 1 describes the process. The input is the binary image X holding the initial location of erythema, that is decided by Eq. (4.19). It returns two sets of pixels, the set of possible scaling pixels L_{scaling} and the set of possible skin

pixels L_{skin}. In the following, we give a detailed description about the dilation, flood fill, and erosion operations used in this algorithm.

Algorithm 1 An algorithm to extract a sample of scaling pixels and a sample of normal skin pixels from an image.

Input: The initial location of the erythema X and image M.

Output: Regions of candidate scaling L_{scaling} pixels and regions of candidate skin L_{skin} pixels.

1: $n \leftarrow 0$
2: **repeat**
3: $X \leftarrow X \oplus U$
4: $n \leftarrow n + 1$
5: **if** an enclosed region is formed in X **then**
6: $X \leftarrow \text{FloodFill}(X)$
7: **end if**
8: **until** no more enclosed regions are formed
9: $L_{\text{Scaling}} \leftarrow M_{x,y} \cap X_{x,y} \ominus nU$
10: $L_{\text{Skin}} \leftarrow M_{x,y} \cap X^{C}_{x,y}$
11: **return** L_{Scaling}, L_{Skin}

The initial value of $X_{x,y}$ acts as a seed for regions of erythema. Each region of erythema is "grown" (the repeat loop in lines 3. to 8.) using dilation operations ($\oplus$) until all regions of scaling are surrounded by regions of erythema. The result is a set of rings where the hole in the centre is likely to contain scaling. For each iteration, the *dilation operation* is expressed as:

$$X \oplus U = \{p_{x,y} | (\hat{U})_{x,y} \cap X \neq \emptyset\} \tag{4.20}$$

where $\emptyset$ is the empty set, U is a structuring element and $p_{x,y}$ is a pixel point at the coordinate (x, y) that has a value of '1' resulting from the dilation operation. $(\hat{U})_{x,y}$ does a reflection operation to the structuring element U and transforms the origin to the point (x, y). In this work, U is chosen to be a disk shape with radius ρ pixels. The dilation expands the erythema regions by examining overlaps between the structuring element and the image. If an overlap happens, the current origin of the structuring element is set to be '1'. In our case, we set $\rho = 5$. An illustration of this structuring element is shown in

Figure 4.7.

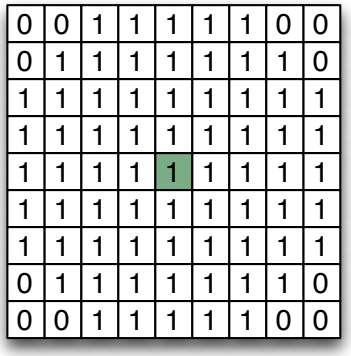


Figure 4.7: A disk-shaped structuring element with 5 pixels radius with the origin marked in green.

After the closed boundary is formed by the dilation operation, a flood fill operation is used to mark the enclosed region with '1'. The *flood fill* operation chooses a point in the enclosed region as a seed and recursively colours the region originating from the seed by examining the connectivity of the region. The recursion stops until the boundary is reached. In this work, 4-connectivity is used due to a high efficiency.

The algorithm erodes ($\ominus$) the erythema regions back to their original width while keeping the closed "doughnut" shape. The *erosion operation* is expressed as:

$$X \ominus nU = \{p_{x,y} | (nU)_{x,y} \subseteq X\} \tag{4.21}$$

where nU is a disk-shaped structure with radius $n\rho$ and $p_{x,y}$ is a pixel point at the coordinate (x, y) that are marked with a value of '1' due to the erosion operation. The erosion operation shrinks the expanded erythema regions. When the structuring element moves in the image, and the structuring element is totally covered in the image, the pixel point of the image that is at the origin of the current structuring element is set to be '1'.

As is shown in Algorithm 1, the variable n is used for counting how many dilation phases have taken place so that we know how many erosion phases need to be performed to regain the original erythema width (line 9). The reason of using the disk-shaped structure with radius $n\rho$ in Eq. (4.21) to regain the original erythema region is that, according to the properties of the dilation and erosion operations, eroding with a structuring ele-

ment n times is equal to doing the erosion with a structuring element whose radius is n times as big as the original one.

$$\begin{aligned} X \ominus \underbrace{U \ldots \ominus}_{n} U &= X \ominus (\underbrace{U \ldots \oplus}_{n} U) \\ &= X \ominus nU \end{aligned} \tag{4.22}$$

Using soft-constrained k-means clustering to choose training sets

The algorithm uses a soft-constrained k-means clustering to select training data from the possible scaling pixel set L_{scaling} and the possible skin pixel set L_{skin}. The constraints for the clustering algorithm are the probabilities that a pixel belongs to one of the clusters. It is assumed that pixels in the potential scaling location have a higher probability of being scaling than normal skin. A similar assumption holds for pixels in the potential skin locations.

The traditional k-means has been adapted by putting such constraints as weights in its clustering process. The initial centroids of the scaling and normal skin clusters are identified by thresholding on the Gabor texture image using Ostu thresholding method. The proposed method is different from the well-known soft-constrained clustering in the dissertation of Wagstaff [139], where the constraints are continuous numbers called the "must-link" and "cannot-link" constraints. In the following, we give a description of the proposed soft-constrained k-means clustering method.

A cluster of scaling pixels C_1 and a cluster of skin pixels C_2 are formed from $L_{\text{scaling}} \cup L_{\text{skin}}$, and then within each of these clusters the pixels with the greater likelihood of being scaling and normal skin respectively are chosen. The feature set F_{scaling} is composed of a scaling contrast map $S_{x,y}$ and a Gabor texture $T_{x,y}$ with the location (x, y) in the regions of L_{scaling} and L_{skin}, so that we have:

$$F_{\text{scaling}} = \{(S_{x,y},\ T_{x,y}) \mid (x, y) \in \text{dom } L_{\text{scaling}} \cup L_{\text{skin}}\} \tag{4.23}$$

where dom means a set of pixels marked as candidate scaling L_{scaling} pixels and candidate

skin $L_{\text{skin}}\}$ pixels.

Let $F_{x,y}$ be an element of F_{scaling} at location (x, y). The objective function for the soft-constrained k-means is defined as the sum of the weighted distances of a sample to each of the cluster centroids:

$$h(\mathcal{C}, \mathcal{O}) = \sum_{i=1}^{2} \sum_{(x,y)\in C_i} W(L_{x,y}, C_i)\|F_{x,y} - \mathcal{O}_i\|^2 \tag{4.24}$$

where $\mathcal{O} = (\mathcal{O}_1, \mathcal{O}_2)$ is the pair of centroids for the clusters $\mathcal{C} = (C_1, C_2)$, and $\|\cdot\|$ is the Euclidean norm. The weight

$$W(L_{x,y}, C_i) = P(L_{x,y}, C_i)^{-1} \tag{4.25}$$

is a weighting function for the location $L_{x,y}$ and class C_i. $P(L_{x,y}, C_i)$ is the probability that a pixel with location $L_{x,y}$ is in the class C_i. This probability is determined by the location of the scaling L_{scaling} and the normal skin L_{skin}. If $L_{x,y}$ is in a region that indicates the same class as C_i, a higher value is assigned, otherwise a lower value is assigned. Note that a further constraint for any single sample at $L_{x,y}$ is that the equality

$$P(L_{x,y}, C_1) + P(L_{x,y}, C_2) = 1 \tag{4.26}$$

must hold.

The objective function h is minimised, when scaling feature samples are assigned to a cluster, to which the weighted distances are smaller than the distances to the other cluster. We have :

$$\{\mathcal{C}, \mathcal{O}\} = \arg\min_{\mathcal{C},\mathcal{O}} h(\mathcal{C}, \mathcal{O}) \tag{4.27}$$

Assuming that the centroids $\mathcal{O}$ and the weighting function $W(L_{x,y}, C_i)$ are independent, to get the minimum of Eq. (4.27) the centroid for each of the two clusters can be obtained by setting the first-order partial derivatives with respect to the centroid to be zero. That is:

$$\frac{\partial h(\mathcal{C}, \mathcal{O})}{\partial \mathcal{O}_i} = 0 \tag{4.28}$$

Resolving the partial derivative function Eq. (4.28), we have:

$$\mathcal{O}_i = \frac{\sum_{(x,y)\in C_i} W(L_{x,y}, C_i) F_{x,y}}{\sum_{(x,y)\in C_i} W(L_{x,y}, C_i)} \tag{4.29}$$

The clustering algorithm partitions the feature set into those that are closer to $\mathcal{O}_1$ and those that are closer to $\mathcal{O}_2$ with respect to the weighted distance. For each partition π_i, we have:

$$\pi_i := \arg\min_{\{\mathcal{C},\mathcal{O}\}} W(L_{x,y}, C_i) \|F_{x,y} - \mathcal{O}_i\|^2 \tag{4.30}$$

The minimum of the objective function $h(\mathcal{C}, \mathcal{O})$ is achieved by iteration of Eq. (4.29) and Eq. (4.30) until the clusters converge.

The training set T_i for the class C_i is taken to be those samples $L_{x,y}$ in the image such that:

$$\frac{W(L_{x,y}, C_i) \|F_{x,y} - \mathcal{O}_i\|^2}{W(L_{x,y}, C_j) \|F_{x,y} - \mathcal{O}_j\|^2} \leq t_h \tag{4.31}$$

where $j = 2$ if $i = 1$ and $j = 1$ if $i = 2$. The threshold t_h is chosen to be 0.1 in order to ensure that training samples have a high likelihood of being within their respective pixel classes. Fig 4.8 shows a diagram of the use of the soft-constrained k-means to select the scaling training set and the skin training set. Only samples inside the circles are those representative samples and are selected as the training samples.

It is noted that the minimisation of $h(\mathcal{C}, \mathcal{O})$ by using the strategy of partial derivatives converges to a local minimum. The minimum of $h(\mathcal{C}, \mathcal{O})$ depends on the initial choice of centroids. A good initialisation is more likely to lead to a global minimum but poorer choices lead to a local minimum. Since the Gabor feature has an advantage in telling scaling from normal skin, it is again used to determine the initial value. The initial centroids for the two classes are chosen to be the average of the following feature vectors:

- $\mathcal{O}_1$ (scaling) is the average of the feature vectors with *high* responses of the Gabor

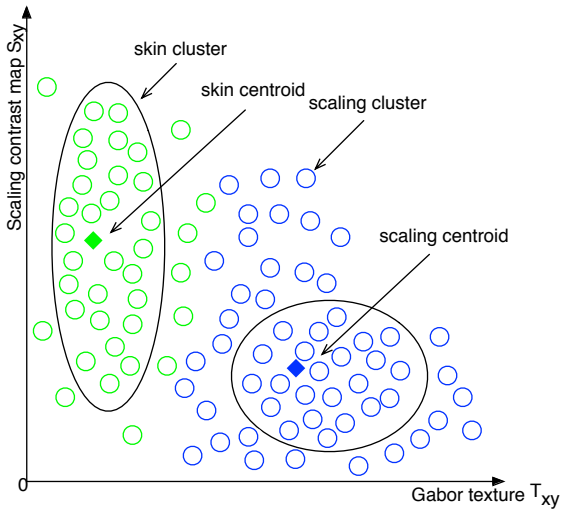


Figure 4.8: Selecting the scaling training set and the skin training set with the soft-constrained k-means clustering. Each circle includes selected feature samples for the corresponding training set, as calculated using Eq. (4.31).

filter for the region L_{scaling}; and

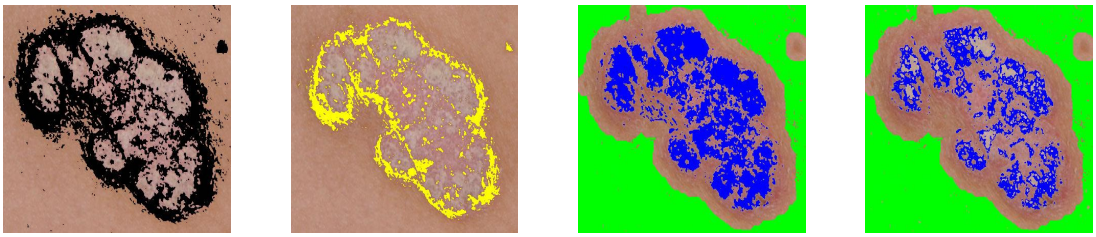
- $\mathcal{O}_2$ (normal skin) is the average of the feature vectors with *low* responses of Gabor filter in the region L_{skin}.

To decide the threshold between high responses of the Gabor filter (the scaling class) and low responses of the Gabor filter (the skin class), the Ostu thresholding method is employed [108]. Ostu's method is based on selecting the lowest point between histograms of two classes. Thus, the intensity variances of the two classes are as small as possible. In this work, we use Ostu's method to work on the histogram of a Gabor texture image to decide the initial centroids of the soft-constrained k-means clustering. The histogram is built by discretising the grayscale intensity of the Gabor texture image into the range of [0,255]. The Ostu thresholding method is introduced below:

The intra-class variance of the scaling class and the skin class with respect to a threshold value at the intensity level x is defined as:

$$\delta^2_{\text{intra}}(x) = \omega_{\text{scaling}}(x)\delta^2_{\text{scaling}}(x) + \omega_{\text{skin}}(x)\delta^2_{\text{skin}}(x) \tag{4.32}$$

where $\delta^2_{\text{scaling}}(x)$ and $\delta^2_{\text{skin}}(x)$ are the total variances of the intensity values in the scaling class and skin class respectively, and $\omega_{\text{scaling}}(x)$ and $\omega_{\text{skin}}(x)$ are the probabilities of being



(a) (b) (c) (d)

Figure 4.9: Stages in the collection of the training sets from the image shown in Figure 4.4. (a) Preprocessing using the scaling contrast filter S where dark coloured pixels are in the matrix M; (b) The approximate localisation of erythema marked in yellow; (c) The candidate regions of scaling (marked in blue) and normal skin marked in green; (d) The representative training sets for scaling marked in blue and normal skin marked in green.

in the scaling class and being in the skin class respectively.

In order to reduce computational complexity, Otsu found that minimising the intra-class variance is equal to maximising the inter-class variance:

$$\begin{aligned} \delta^2_{\text{inter}}(x) &= \delta^2 - \delta^2_{\text{intra}}(x) \\ &= \omega_{\text{scaling}}(x)\omega_{skin}(x)(\mu_{\text{scaling}}(x) - \mu_{\text{skin}}(x))^2 \end{aligned} \tag{4.33}$$

where δ^2 is the total variance of the histogram. $\mu_{\text{scaling}}(x)$ and $\mu_{\text{skin}}(x)$ are mean values of the respective classes with association to the current thresholding value x.

The calculation of the variance in Eq. (4.32) is replaced with the simple calculation of mean difference in Eq. (4.33). The optimal threshold value t_{Gabor} is the one corresponding to the maximum of inter-class variance:

$$t_{\text{Gabor}} = \arg\max_x \delta^2_{\text{inter}}(x) \tag{4.34}$$

The maximum is achieved by enumerating the intensity level in the range. Figure 4.9 visually illustrates the process of collecting training sets of scaling and normal skin from the skin image. The training sets are collected from non-black pixels in Figure 4.9(a). After using the approximate location of erythema in Figure 4.9(b), the rough locations of scaling and normal skin are identified. It is observed in Fig 4.9(d) that the selective training sets for scaling and normal skin are composed of representative pixel samples.

4.3.3 The SVM-based MRF to identify scaling pixels

After the training sets for scaling and normal skin are identified in each image, an SVM-based MRF model is proposed to classify scaling pixels from normal skin pixels in the image.

A pixel-based classification method can resolve a large range of problems, even when scaling and normal skin occur at psoriasis lesion boundaries. However, the classification-based segmentation more often depends on the image structure and the neighbourhood of the pixel being classified than on clear distinctions in a feature space. An MRF is formulated precisely with this type of problem in mind.

An MRF is a graphical model for a random labelling process, where the interaction is only considered between two neighbouring sites. In an image segmentation problem, this kind of an interaction describes the homogeneous property of a segmented region quite well. The MRF works together with Maximum a Posteriori (MAP) estimation to identify homogeneous regions. It aims to maximise the probability of pixel labelling given the image feature and defined graph structure [140].

In this work, let

$$A = \{(S_{x,y},\ T_{x,y})|\ M_{x,y} = 1\} \tag{4.35}$$

be the set of scaling features for all of the image pixels that are thresholded as non-erythema pixels in Eq. (4.18). The segmentation ω is considered as an MRF. It is aimed to label the scaling and normal skin on the rest of the image coordinate, where erythema pixels and other darkened pixels are thresholded out. We have:

$$\omega = \{\omega_{x,y}|(x,y) \in \mathcal{S}\} \tag{4.36}$$

where $\omega_{x,y}$ is a labelling output, or a classification result for a pixel at the coordinate (x,y), $\omega_{x,y} \in \{\text{scaling, normal skin}\}$, and $\mathcal{S}$ is the set of sites in the MRF. In this work, $\mathcal{S}$ are composed of pixels that are not thresholded out in the binary image M, $\mathcal{S} = \{(x,y)|M_{x,y} = 1\}$.

In particular, we define the likelihood term of the MRF by using the distance of a feature sample to the hyperplane trained by the SVM. By delivering such a definition,

we describe the likelihood term though already classified results and avoid the parameter estimation of a Gaussian distribution that is normally used to model the likelihood term. Additionally, the smoothness term of the MRF is modelled by using the Euclidean distance of scaling features. Thus, homogeneous regions are modelled with the consideration of feature differences between the current pixel and its neighbouring pixels. This defines a homogeneous region well, where neighbouring pixels have similar values of the scaling features.

The MRF modelling is resolved by using a graph-cut method. For this binary classification problem, the graph-cut method introduces two additional vertices as two labels to the graph of the MRF, and constructs edge weights to conduct the labelling. The graph cut method has a higher efficiency than the traditional simulated annealing and iterated conditional modes methods [141, 142].

Constructing the MRF-MAP framework for scaling segmentation

An MRF labelling process is about the labelling for the whole graph. The Markovian property is satisfied, if the probability of labelling the current pixel is only related to the labelling of the pixels in its neighbourhood, and is not affected by the labelling of pixel points outside of the neighbourhood:

$$P(\omega_{x,y}|\omega_{\mathcal{S}-(x,y)}) = P(\omega_{x,y}|\omega_{N_{x,y}}) \tag{4.37}$$

where $N_{x,y}$ is a neighbourhood of the pixel at coordinate (x, y).

According to the Hammersley-Clifford theorem, a Gibbs random field is equivalent to an MRF [143]. It gives a global description about the MRF labelling process, and has the definition of:

$$P(\omega) = \frac{1}{Z}\exp(-U(\omega)) \tag{4.38}$$

where Z is a normalising constant, and $U(\omega)$ is the energy function, that is the summation of clique potentials over a set of cliques.

When it comes to the segmentation problem, we only use pairwise neighbourhood

cliques to describe homogeneity of a region. The probability of labelling for the whole image $P(\omega)$ is rewritten as:

$$P(\omega) = \frac{1}{Z}\exp(-\sum_{x,y}\sum_{x',y'\in N_{x,y}} V(\omega_{x,y}, \omega_{x',y'})) \tag{4.39}$$

where $N_{x,y}$ is an 8-connected neighbourhood, (x', y') is the coordinate of a pixel in the neighbourhood $N_{x,y}$, and the function $V(\omega_{x,y}, \omega_{x',y'})$ is a clique potential function, that gives a description of homogeneity of two neighbouring pixels and is called a smoothness term as well.

The maximum a posteriori (MAP) estimation theory provides a solution to resolve the segmentation problem based on the MRF modelling. The segmentation ω can be considered as an estimation of pixel labelling given the feature set A. It can be achieved through a maximisation of the posterior.

$$\omega = \text{argmax} P(\omega|A) \tag{4.40}$$

According to the bayesian theory, we have:

$$P(\omega|A) = \frac{P(A|\omega)P(\omega)}{P(A)} \tag{4.41}$$

Since the feature set A is known, the probability of the feature set $P(A)$ is a constant. Eq. (4.40) can be rewritten by using the likelihood term $P(A|\omega)$ and the prior $P(\omega)$:

$$\omega = \text{argmax} P(A|\omega)P(\omega) \tag{4.42}$$

The likelihood term $P(A|\omega)$ can be expressed through the likelihood of each variable, given the assumption that feature variables are conditionally independent with the given class labels.

$$P(A|\omega) = \prod_{x,y} P(A_{x,y}|\omega_{x,y}) \tag{4.43}$$

Consequently, by using Eq. (4.39) and Eq. (4.43), the MAP maximisation problem defined in Eq. (4.42) is converted into a minimisation problem through a minus logarithm

operation. That is:

$$\omega = \text{argmin} \sum_{x,y} -\ln P(A_{x,y}|\omega_{x,y}) + \sum_{x,y} \sum_{x',y' \in N_{x,y}} V(\omega_{x,y}, \omega_{x',y'}) \tag{4.44}$$

The minimised term given by Eq. (4.44) is also called the cost function of the MRF model. Therefore, defining an MRF is based on the definition of individual likelihood term $P(A_{x,y}|\omega_{x,y})$ and the smoothness term $V(\omega_{x,y}, \omega_{x',y'})$. The detailed definitions of the two terms are described in the following section.

Defining the MRF based on the SVM classification

In this work, the likelihood term is modelled by considering the distance of a feature sample to the hyperplane of the SVM (see Section 3.4 for an introduction of the SVM method). The assumption here is that the probability of a certain feature point given a particular class label can be calculated from its distance and direction to the hyperplane. If the feature is close to the hyperplane then the pixel may have been misclassified, while if it is far from the hyperplane then it is assumed that the probability of the pixel being correctly classified is high.

Specifically, for each item in the likelihood term of the MRF, we have:

$$P(A_{x,y}|\omega_{x,y}) = \begin{cases} \frac{1}{1+\exp(-d(A_{x,y}))} & \text{if } a_{x,y} = \omega_{x,y} \\ \frac{\exp(-d(A_{x,y}))}{1+\exp(-d(A_{x,y}))} & \text{otherwise} \end{cases} \tag{4.45}$$

where $a_{x,y}$ is the classification result from the SVM, and $d(A_{x,y})$ is a measurement that is linearly related with the distance of $A_{x,y}$ to the SVM hyperplane. The definition of $d(A_{x,y})$ is given by:

$$\begin{aligned} d(A_{x,y}) &= |\mathbf{w} \cdot \phi(A_{x,y}) + b| \\ &= |\sum_{x',y'} a_{x',y'} \lambda_{x',y'} \phi(F_{x',y'}) \cdot \phi(A_{x,y}) + b| \\ &= |\sum_{x',y'} a_{x',y'} \lambda_{x',y'} K(F_{x',y'}, A_{x,y}) + b| \end{aligned} \tag{4.46}$$

where $\lambda_{x',y'}$ is a Lagrange multiplier [144], b is the intercept of the hyperplane, $\phi(\cdot)$ is the mapping function, and $K(\cdot,\cdot)$ is the kernel function. In this work, a Gaussian Radial Basis function is used because of its simplicity and computational efficiency.

This distribution assigns a higher probability to $P(A_{x,y}|\omega_{x,y})$ if the class label for $\omega_{x,y}$ is the same as that given by the SVM. Moreover, the probability for label $\omega_{x,y}$ is higher if the feature point is further away from the hyperplane. Conversely, if it is in a different class to that given by the SVM, $P(A_{x,y}|\omega_{x,y})$ is assigned a lower probability, and the further it is away from the hyperplane, the lower $P(A_{x,y}|\omega_{x,y})$ is.

The smoothness term of the MRF $V(\omega_{x,y},\omega_{x',y'})$ is modelled by considering the feature values and labels of the current and neighbouring pixels. It describes the homogeneity of a region. In a homogeneous region, pixels are connected with each other and have similar feature values. If there is a larger disparity between the feature values of two neighbouring pixels, the two pixels are less likely to be segmented as a region. The smoothness term $V(\omega_{x,y},\omega_{x',y'})$ is modelled as:

$$V(\omega_{x,y},\omega_{x',y'}) = \begin{cases} k \cdot \exp(\|A_{x,y} - A_{x',y'}\|) & \text{if } \omega_{x,y} \neq \omega_{x',y'} \\ 0 & \text{otherwise} \end{cases} \tag{4.47}$$

where k is a penalty constant. $V(\omega_{x,y},\omega_{x',y'})$ gives the probability of observing the features of the pixel $A_{x,y}$ given the features at a neighbouring pixel $A_{x',y'}$. The probability is higher for homogeneous regions and lower for heterogenous regions. Furthermore, higher penalty values are assigned if neighbouring pixels with different labels show larger difference of the feature values.

Resolving the MRF modelling using a graph-cut method

Since the cost function of the MRF model is submodular, a graph cut method is used to give a binary labelling of the vertices in the MRF. It is proved in [145] that the minimised solution of the MRF cost function is equivalent to the minimum cut cost found on a graph, that is usually solved through a maximum flow algorithm. In the following, we give a brief description about how to solve the binary MRF labelling problem through a graph-cut method.

In a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set of vertices including the source l_s and the terminal sink l_t, and $\mathcal{E}$ is the set of edges with non-negative weights. A cut is a set of edges that separate the graph into two sub-graphs, and disconnect the source l_s and the sink l_t. A cut cost is the summation of the weights of the cut edges. When it comes to the binary-label MRF problem, besides the source l_s and the terminal sink l_t, the vertex set $\mathcal{V}$ includes the pixel points that are defined as vertices in the MRF. Pixels connected with the source l_s are assigned with one label l_1, and pixels connected with the sink l_t are assigned with the other label l_2.

In the binary-label MRF model, let $V_i(l_i)$ be the individual negative log likelihood term of a pixel point i with the label l_i, and $V_{i,j}(l_i, l_j)$ be the smoothness term between the pixel point i with the label l_i and the pixel point j with the label l_j. If the submodularity condition holds, we have:

$$V_{i,j}(l_1, l_2) + V_{i,j}(l_2, l_1) \geq V_{i,j}(l_1, l_1) + V_{i,j}(l_2, l_2) \tag{4.48}$$

There are two parts for building the graph, in which the graph cut is equivalent to the cost function of the MRF model. The two parts are explained below:

- For the likelihood term, if $V_i(l_1) > V_i(l_2)$, add an edge from l_s to i with the weight $V_i(l_1) - V_i(l_2)$; otherwise add an edge from i to l_t with the weight $V_i(l_2) - V_i(l_1)$.
- To represent the smoothness term, three edges are added to the graph:
    1. If $V_{i,j}(l_1, l_2) > V_{i,j}(l_1, l_1)$, add an edge from l_s to i with the weight $V_{i,j}(l_1, l_2) - V_{i,j}(l_1, l_1)$; otherwise add an edge from i to l_t with the weight $V_{i,j}(l_1, l_1) - V_{i,j}(l_1, l_2)$.
    2. If $V_{i,j}(l_1, l_2) > V_{i,j}(l_2, l_2)$, add an edge from j to l_t with the weight $V_{i,j}(l_1, l_2) - V_{i,j}(l_2, l_2)$; otherwise add an edge from l_s to j with the weight $V_{i,j}(l_2, l_2) - V_{i,j}(l_1, l_2)$.
    3. Add an edge from the pixel point i to j, with the weight $V_{i,j}(l_1, l_2) + V_{i,j}(l_2, l_1) - V_{i,j}(l_1, l_1) - V_{i,j}(l_2, l_2)$.

The minimum cut of the constructed graph is solved through its dual form: a maximum flow algorithm. In the maximum flow algorithm, edge weights act like the water

capacities of pipes. The maximum flow is the largest flow of water from l_s to l_t. When there is a maximum flow from l_s to l_t, edges for the minimal cut cost are saturated edges that are pipes at capacity in the maximum flow problem. In this work, the max-flow algorithm proposed in [146] is applied. This algorithm iteratively searches a path from the source l_s to the sink l_t and augments the path so that some edges are at capacity.

Figure 4.10 demonstrates the MRF for segmenting scaling based on the SVM. Scaling pixels that are misclassified by the SVM as normal skin pixels (see Figure 4.10(a)) are eliminated by the MRF in the final result (see Figure 4.10(c)).

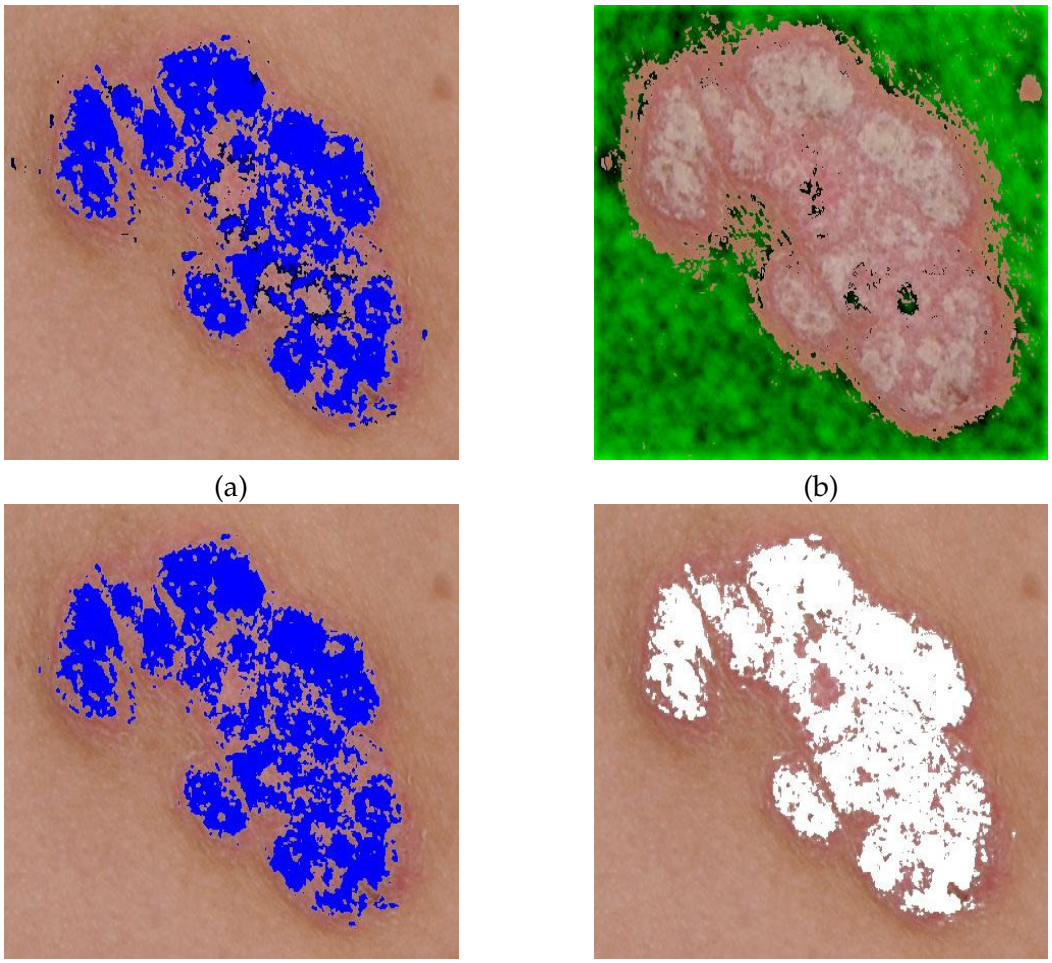
4.4 Experimental results: scaling segmentation

The algorithm has been tested on a set of 103 images, which are collected from a dataset containing 722 psoriasis scaling images. The images were chosen so that there was a good distribution of images taken under different lighting conditions and at different angles, images with shadows, images with wrinkles, and images with hair. The images in each category were randomly selected.

The sizes of the images vary from 940×666 to 161×142 and the total set covers plaque psoriasis, erythrodermic psoriasis, guttate psoriasis, and pustular psoriasis. The images are captured with high resolution digital cameras in an indoor environment under stable illumination provided by incandescent lamps, among which 40 images are captured by a Fuji Pix S2, 54 images are captured by a Nikon D300 and 9 images are captured by a Nikon D3100. The skin type varies with people from Asian or Caucasian background.

We set the threshold $t_s = 0.004$ for the scaling contrast map in the definition of $M_{x,y}$ (see Eq. (4.18)) as a balance between removing erythema and retaining scaling. Note that even potential regions of scaling are highly likely to contain areas of normal skin and so the parameter values are chosen to reflect this fact. The parameters used in the soft-constrained k-means procedure are given in Table 4.2. The penalty constant in SVM-based MRF classification is set as $k = 4$.

In this section, three series of experiments are conducted. The first series of experi-



(a) (b)

(c) (d)

Figure 4.10: The segmentation of scaling for the lesion in Figure 4.4. (a) The distance of a scaling pixel to the SVM hyperplane where darker pixels are closer to the SVM hyperplane; (b) The distance of normal skin pixels to the SVM hyperplane where darker green pixels are closer to the SVM hyperplane; (c) SVM-based MRF localisation of scaling marked in blue; (d) Ground truth of scaling marked in white colour.

ment compares the soft-constrained k-means with the k-means and fuzzy c-means clustering methods by evaluating the training sets collected using the three methods; the second series of experiment compares the performance of scaling segmentation using our method with an SVM and an MRF; the third series of experiment evaluates parameters used in the algorithm: the threshold value t_s used for removing erythema and other dar pixels in Eq. (4.18), the probability $P(L_{x,y}, C_i)$ used for the definition of weights in the soft-constrained k-means (see Eq. (4.25)), and the penalty constant k used in the SVM-based

Parameter	Value
$P\{L_{x,y} \in L_{\text{scaling}}, C_1\}$	0.67
$P\{L_{x,y} \in L_{\text{scaling}}, C_2\}$	0.33
$P\{L_{x,y} \in L_{\text{skin}}, C_1\}$	0.20
$P\{L_{x,y} \in L_{\text{skin}}, C_2\}$	0.80

Table 4.2: Parameters in the soft-constrained k-means procedure.

MRF classification (see Eq. (4.47)).

4.4.1 Evaluating the Training Sets Obtained with the Constrained K-means Clustering

Performance of the soft-constrained k-means is evaluated using Mean Absolute Difference (MAD) [147] and Spatial Support (SS) [148], that provide assessments of consistency between training samples and the corresponding scaling and skin pixels in the actual image. The MAD and the SS measurements are defined as:

$$\begin{aligned} \text{MAD}_\text{m} &= \frac{|Q_m| - |Q_m \bigcap G_m|}{|Q_m|} \\ \text{SS} &= \frac{1}{2} \sum_{m=1}^{2} \frac{|Q_m \bigcap G_m|}{|Q_m \bigcup G_m|} \end{aligned} \tag{4.49}$$

where Q_m is the set of pixels in the cluster m and G_m is the corresponding ground truth, that is, the set of pixels that are the true scaling pixels or the true skin pixels. MAD is able to separately measure clustering results for skin and scaling, while SS is an overall assessment.

MAD measures the degree of difference between the clustering results and the ground truth. SS is calculated as the mean percentage of the clustering results that overlap the "ground truth". The two measures indicate the degree to which the cluster is faithful to the "ground truth". When MAD values are closer to 0 and SS values are closer to 1, the training sets are more consistent with the "ground truth", and will have a better performance.

The performance of the soft-constrained k-means is compared with a traditional k-

means and with a fuzzy c-means. The traditional k-means and the fuzzy c-means are chosen because both methods have proved very popular in the clustering literature [149, 150]. We choose the same initial centroids for the k-means and the fuzzy c-means as explained in Section 4.3.2. In the traditional k-means, no weights are added to the distance between a sample point and a clustering centroid. The objective function of the traditional k-means is:

$$h_{Kmeans}(\mathcal{C}, \mathcal{O}) = \sum_{i=1}^{2} \sum_{(x,y)\in C_i} \|F_{x,y} - \mathcal{O}_i\|^2 \tag{4.50}$$

A fuzzy c-means classifier is derived by incorporating "fuzzy" concepts into the clustering process [151]. Weights are assigned to the distances to centroids and change with the distances. When a sample point is closer to a clustering centroid, the associated weight is assigned with a higher value. And when a sample point is further from a centroid, the associated weight becomes lower. A standard fuzzy c-means objective function is expressed as below:

$$h_{FCM}(\mathcal{C}, \mathcal{O}) = \sum_{i=1}^{2} \sum_{(x,y)\in C_i} u_{(x,y),i}^{p_f} \|F_{x,y} - \mathcal{O}_i\|^2 \tag{4.51}$$

where $u_{(x,y),i}$ is the degree of membership of a feature sample $F_{x,y}$ in the cluster C_i, which subjects to the constraints: $\sum_{i=1}^{2} u_{(x,y),i} = 1$ and $u_{(x,y),i} \in [0, 1]$, and p_f is a weighting exponent. The weighting exponent p_f determines the amount of fuzziness of the resulting classification. As pointed out in [152], the optimal choice of the weighting exponent is in the range of [1,5, 2.5]. In this experiment, we chose the midpoint and have $p_f = 2$.

We again use Eq. 4.31 to filter the clusters to identify training sets for scaling and skin. A part of the clustering results is presented in Figure 4.11. We can see that the soft-constrained k-means localised many more training samples than either the k-means or the Fuzzy C means. Moreover, the soft-constrained k-means is more robust to changes in skin colour. When the skin is not evenly illuminated, the clusters obtained with the k-means and the fuzzy c-means do not correspond well to the scaling and skin pixels in the image. When the skin colour is similar to the scaling colour, the centroids initially identified are swapped by the fuzzy c-means algorithm because of the effect of its membership

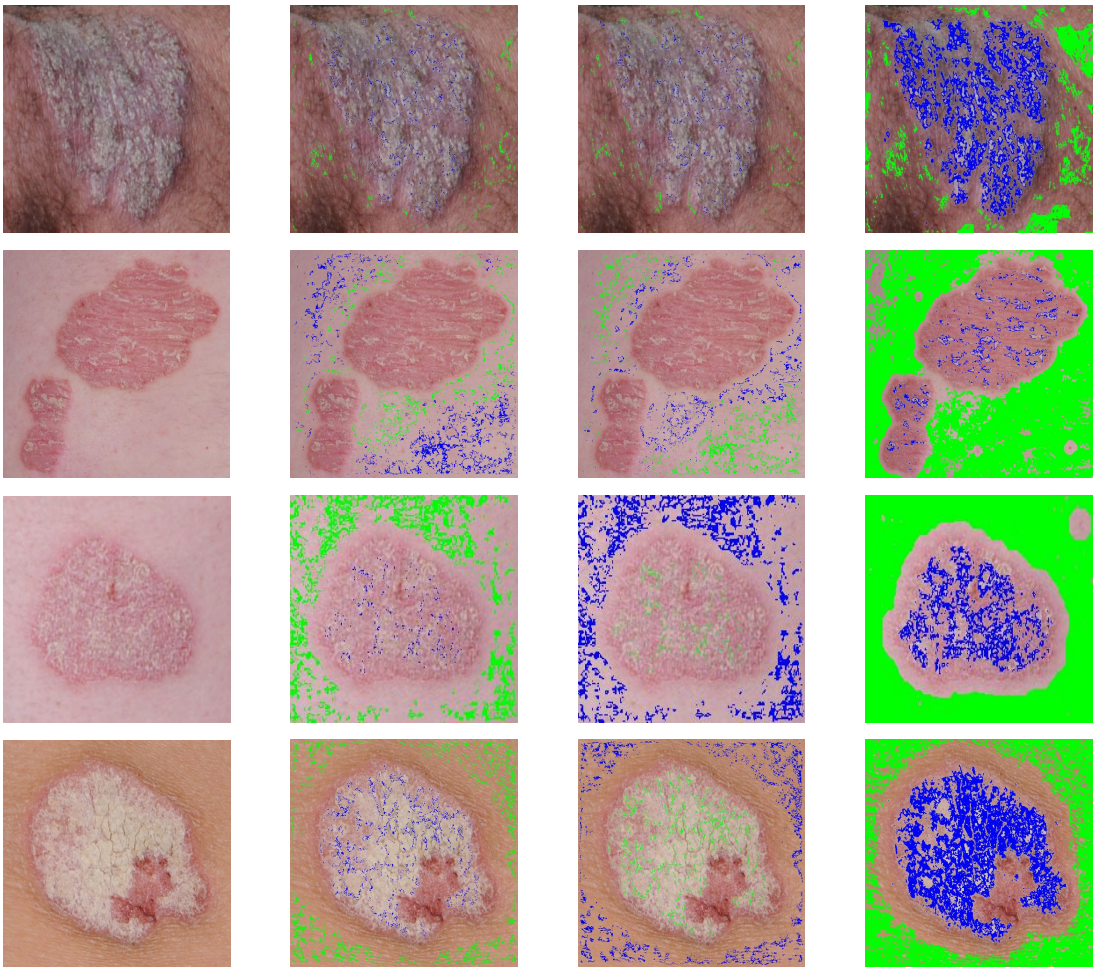


Figure 4.11: Clustering results. Training samples for scaling are marked in blue and training samples for skin are marked in green. (a) Original image; (b) Training sets from the k-means; (c) Training sets from the Fuzzy C-means; (d) Training sets from the soft-constrained k-means.

function.

Performances of the three clustering methods are shown in Table 4.3. For skin, the soft-constrained k-means has a better MAD value than the traditional k-means and the fuzzy c-means. For scaling, the soft-constrained k-means shows an obvious advantage to the Fuzzy C-means in their MAD, but a slight inferiority to the k-means. Moreover, the soft constraints k-means has a much better SS over both the skin and scaling clusters.

Table 4.3: Clustering method comparison.

	MAD of skin	MAD of scaling	SS
k-means	0.0231	0.6045	0.0889
fuzzy c-means	0.1873	0.8846	0.0220
Soft-constrained k-means	0.0110	0.6092	0.4598

4.4.2 Evaluating the Localisation of Scaling

A quantitative evaluation of the segmentation results is made by measuring *sensitivity*, *specificity* and *dice*. The calculations of sensitivity and specificity are described in Section 3.5.2. Sensitivity measures the ability of the classifier to identify "true positives", which in our case, is the percentage of ground truth pixels labelled as scaling pixels by the classifier. Specificity measures the "true negative", which in our case is the percentage of non-scaling classified by the algorithm.

Dice's coefficient is proposed in [153]. It measures the overlap between the segmentation region and the "ground truth" region by evaluating the similarity of the two regions. For this binary segmentation problem the Dice's coefficient is an overall evaluation for the accuracy of scaling segmentation and the associated accuracy of normal skin segmentation. It is defined as:

$$\text{Dice} = \frac{2|L_S \bigcap G_S|}{|L_S| + |G_S|}$$

where L_S is the scaling segmentation result and G_S is the associated "ground truth".

In order to evaluate the significance of the measurements of sensitivity, specificity and dice, a bootstrap analysis is used. The bootstrap assumes that the true measurement can be estimated by resampling a group of sample sets from the given sample set. The resampled sample set has the same number of samples as the given sample set, and are treated as sample sets from the true senario. The true error of a measurement is estimated to be the standard variance of the measurement derived from the resampled samples [154]. In this research, we resampled 1000 sample sets for the bootstrap analysis.

We compare the performance of the proposed scaling classifier with an SVM alone and a standard MRF alone as in [140]. In [140], the MRF is modelled to segment colour textured images. It is constructed with the likelihood term of a Gaussian Distribution

and with the smoothing term $V(\omega_{x,y}, \omega_{x',y'})$ defined by a Gibbs distribution. In the experiment, we model the Gaussian Distribution using the feature set A. The likelihood term is expressed as below:

$$P(A_{x,y}|\omega_{x,y}) = \frac{1}{\sqrt{4\pi^2|\delta_{\omega_{x,y}}|}}\exp(-\frac{1}{2}(A_{x,y} - \mu_{\omega_{x,y}})^t \delta_{\omega_{x,y}}(A_{x,y} - \mu_{\omega_{x,y}})) \tag{4.52}$$

where $\delta_{\omega_{x,y}}$ and $\mu_{\omega_{x,y}}$ are the variance and the mean of a class $\omega_{x,y}$ respectively. They are derived from the associated training sets.

The smoothing term with a Gibbs distribution is expressed as:

$$V(\omega_{x,y}, \omega_{x',y'}) = \begin{cases} 1 & \text{if } \omega_{x,y} \neq \omega_{x',y'} \\ -1 & \text{otherwise} \end{cases} \tag{4.53}$$

Some examples of the segmentation of scaling are shown in Figure 4.12. It is observed that our segmentation results are better than the SVM and the MRF in the aspect of differentiating normal skin from scaling when the normal skin is around psoriatic lesions. As illustrated by the lesions in the second and third rows of Figure 4.12, the SVM and the MRF misclassify normal skin around psoriatic lesions as psoriasis, but our combined Markov Random Field and SVM classifier corrects the misclassification.

To assess the performance of the training sets collected in Section 4.3.2, a comparison of the classification results is made with training sets collected manually. Examples of segmentation results with manually collected training set are shown in Figure 4.13. The results indicate that our method is still superior to the SVM and the MRF when training samples are manually collected.

Tables 4.4 and 4.5 respectively show the results when training sets are obtained with the soft-constrained k-means, and when training sets are collected manually for all 103 images. It is found that the sensitivity evaluation of the proposed method is less clear cut in Tables 4.4 and 4.5. Though the proposed method does not show an advantage in either automatic training set collection or manual training set collection, the sensitivity of our method is very similar to the SVM when training sets are obtained using the

soft-constrained k-means. The oversegmentation of the MRF method with the Gaussian likelihood function, where some normal skin is also classified as scaling, causes the sensitivity to be the highest in both cases. As stated in Section 3.5.2, the sensitivity is more important in dermatology area, but sensitivity alone is not sufficient to differentiate between the classifiers in this case. Therefore a combination of Dice and specificity is used.

The differences are in the specificity results. The specificity for both automatic and manually selected training sets is higher for the combined classifier than either SVM or MRF individually, indicating fewer false non-scaling pixels classified as scaling pixels.

Moreover Dice's coefficient is much higher for the combined classifier than either the SVM or the MRF indicating higher similarity between the sets of scaling pixels and non-scaling pixels as found by the combined classifier and the "ground truth".

The comparison of classification results from the training sets selected by the soft-constrained k-means and the manually selected sets indicates a difference as well. The manually selected training set has a higher specificity and dice for the MRF and has a higher dice for the SVM. Additionally, for the proposed method the difference between the two groups of training sets in the specificity and the dice is not very big. When the measures do not show a clear advantage for one method over another in terms of results then the automated method where human intervention is not required is to be preferred.

The robustness of the algorithm is tested against 16 images with wrinkled skin, 29 images with hair, 19 images with shadows, 11 images where the imaging direction has been changed, and 11 image where the illumination is changed. Some examples of the results are shown in Figure 4.14. The evaluation of the images is summarised in Table 4.6. The evaluation also considered training sets collected manually and training sets collected using the soft-constrained k-means for both the SVM and MRF evaluation.

The results in Table 4.6 show the robustness of the proposed algorithm against shadows, hair, wrinkles and changes of imaging environment. Recall from Section 3.5.2, the sensitivity plays an important role compared withe the specificity. However, in this case, since the difference of the sensitivity values between the classifiers is not obvious, the evaluation of using the sensitivity alone is not enough to illustrate which classifier is advantageous. The dice measurement for our method is always higher than the other two

Table 4.4: A comparison of scaling segmentation results with training sets from the soft-constrained k-means.

	Sensitivity	Specificity	Dice
SVM	$0.7303 \pm 3e-4$	$0.8764 \pm 1e-4$	$0.3817 \pm 3e-4$
MRF	$0.7638 \pm 3e-4$	$0.8677 \pm 9e-5$	$0.3642 \pm 3e-4$
Proposed method	$0.7229 \pm 3e-4$	$0.8946 \pm 8e-5$	$0.4249 \pm 3e-4$

Table 4.5: A comparison of scaling segmentation results using manually selected training sets.

	Sensitivity	Specificity	Dice
SVM	$0.7855 \pm 4e-4$	$0.8779 \pm 9e-5$	$0.3777 \pm 3e-4$
MRF	$0.7810 \pm 4e-4$	$0.8210 \pm 9e-5$	$0.3304 \pm 3e-4$
Proposed method	$0.7667 \pm 3e-4$	$0.9023 \pm 8e-5$	$0.4349 \pm 3e-4$

classifiers where manually selected training sets are used. In all cases, either sensitivity or specificity for our method is highest as well. The use of the contrast map enables our algorithm to differentiate scaling from shadows, images captured in high illumination and images captured in low illumination. The changes of imaging direction do not affect the segmentation results, even though the lighting condition changes in this situation. In addition, our algorithm shows robustness to wrinkles and skin with short hair. This is the contribution of the Gabor features. The bank of Gabor filters developed in Section 4.2.2 is good at characterising the difference between scaling and wrinkles as well as short hair, due to the use of multiple scales and orientations. However, when the hair is long and clear in the image, our Gabor features fail to suppress the disturbance.

4.4.3 Parameter Estimation

Three groups of parameters need to be estimated. They are the threshold value t_s used for removing erythema and other dark pixels in Eq. (4.18), the probability $P(L_{x,y}, C_i)$ used in the soft-constrained k-means clustering in Eq. (4.25), and the penalty constant k from SVM-based MRF classification in Eq. (4.47).

The threshold value t_s is plays a major role in converting the problem into a binary classification problem that is only separating scaling from normal skin. Sensitivity values for removing erythema and the remaining scaling are calculated (as shown in Figure (4.15)). When $t_s = 0$, the sensitivity for removing erythema is 0.6477, while the

Table 4.6: A comparison of scaling segmentation results for images under a variety of different conditions using manually selected training sets.

	Sensitivity	Specificity	Dice
Images with shadows			
SVM	$0.8048 \pm 8e-4$	$0.8734 \pm 2e-4$	$0.4265 \pm 7e-4$
MRF	$0.8334 \pm 9e-4$	$0.8502 \pm 2e-4$	$0.3995 \pm 7e-4$
Proposed method	$0.8179 \pm 9e-4$	$0.8707 \pm 2e-4$	$0.4505 \pm 7e-4$
Images with wrinkles			
SVM	$0.7303 \pm 6e-4$	$0.8347 \pm 3e-4$	$0.5180 \pm 6e-4$
MRF	$0.7653 \pm 6e-4$	$0.8019 \pm 3e-4$	$0.5010 \pm 5e-4$
Proposed method	$0.7278 \pm 5e-4$	$0.8849 \pm 2e-4$	$0.5503 \pm 5e-4$
Images with hair			
SVM	$0.7334 \pm 6e-4$	$0.9009 \pm 2e-4$	$0.4427 \pm 5e-4$
MRF	$0.7273 \pm 6e-4$	$0.8426 \pm 2e-4$	$0.3954 \pm 5e-4$
Proposed method	$0.7591 \pm 5e-4$	$0.8769 \pm 1e-4$	$0.4737 \pm 5e-4$
Images about changes of imaging direction			
SVM	0.8264 ± 0.0014	$0.8255 \pm 4e-4$	0.3747 ± 0.0012
MRF	0.7176 ± 0.0012	$0.6444 \pm 4e-4$	0.2351 ± 0.0011
Proposed method	0.8035 ± 0.0014	$0.8855 \pm 3e-4$	0.4398 ± 0.0012
Images about changes of illuminance			
SVM	0.7948 ± 0.0011	$0.8225 \pm 3e-4$	$0.3291 \pm 8e-4$
MRF	$0.7370 \pm 9e-4$	$0.8271 \pm 3e-4$	$0.2739 \pm 8e-4$
Proposed method	0.7830 ± 0.0011	$0.9112 \pm 2e-4$	$0.4001 \pm 8e-4$

sensitivity for remaining scaling is 0.8122. Even though a low threshold value obtains a high sensitivity for scaling, it prevents erythema from being separated effectively in the training sample collection stage. The value used in the evaluation was $t_s = 0.004$.

The probability $P(L_{x,y} \in L_{\text{scaling}}, C_i)$ in the soft-constrained k-means is the complement of $P(L_{x,y} \in L_{\text{skin}}, C_i)$. Two parameters $P(L_{x,y} \in L_{\text{scaling}}, C_1)$ and $P(L_{x,y} \in L_{\text{skin}}, C_2)$ are examined, since the other two are calculated from these . The average SS is shown in Figure 4.16 and is derived by calculating mean values over a range of probabilities of the other potential regions. The highest SS for $P(L_{x,y} \in L_{\text{scaling}}, C_1)$ and $P(L_{x,y} \in L_{\text{skin}}, C_2)$ are achieved at 0.1 and 0.9 respectively. In the experiment, $P(L_{x,y} \in L_{\text{scaling}}, C_1) = 0.33$ and $P(L_{x,y} \in L_{\text{skin}}, C_2) = 0.8$ are approximations of the probabilities. They are very close to the associated highest value.

The results of varying the penalty constant k in the SVM-based MRF classifier is illustrated in Figure 4.17. Observe that the accuracy of the segmentation results changes with the penalty constant k. When k is too small, for example when $k = 0$, the algorithm does

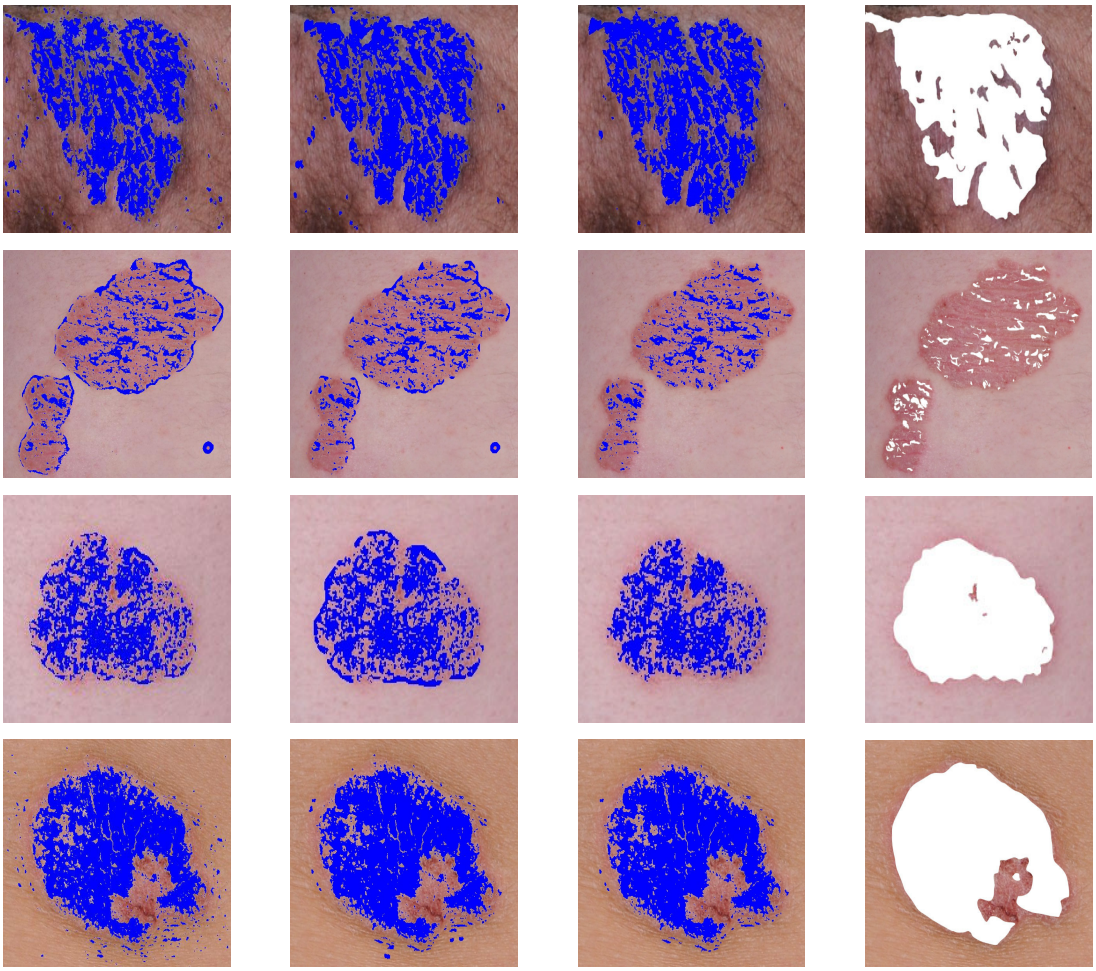


Figure 4.12: Classification results of the original images in Figure 4.11 with training sets from the soft-constrained k-means. Detected scaling is marked in blue. (a) SVM segmentation; (b) MRF segmentation; (c) Our segmentation; (d) Ground truth.

not work for normal skin pixels around the lesion. Increasing k increases the influence from neighbouring pixels' labelling. The results become stable when $k \geq 4$.

4.5 Summary

A semi-automatic classification algorithm is presented in this chapter to segment scaling from psoriasis skin images. Firstly, in the feature analysis stage, it begins with the scaling contrast map construction followed by the Gabor texture analysis. Secondly, the

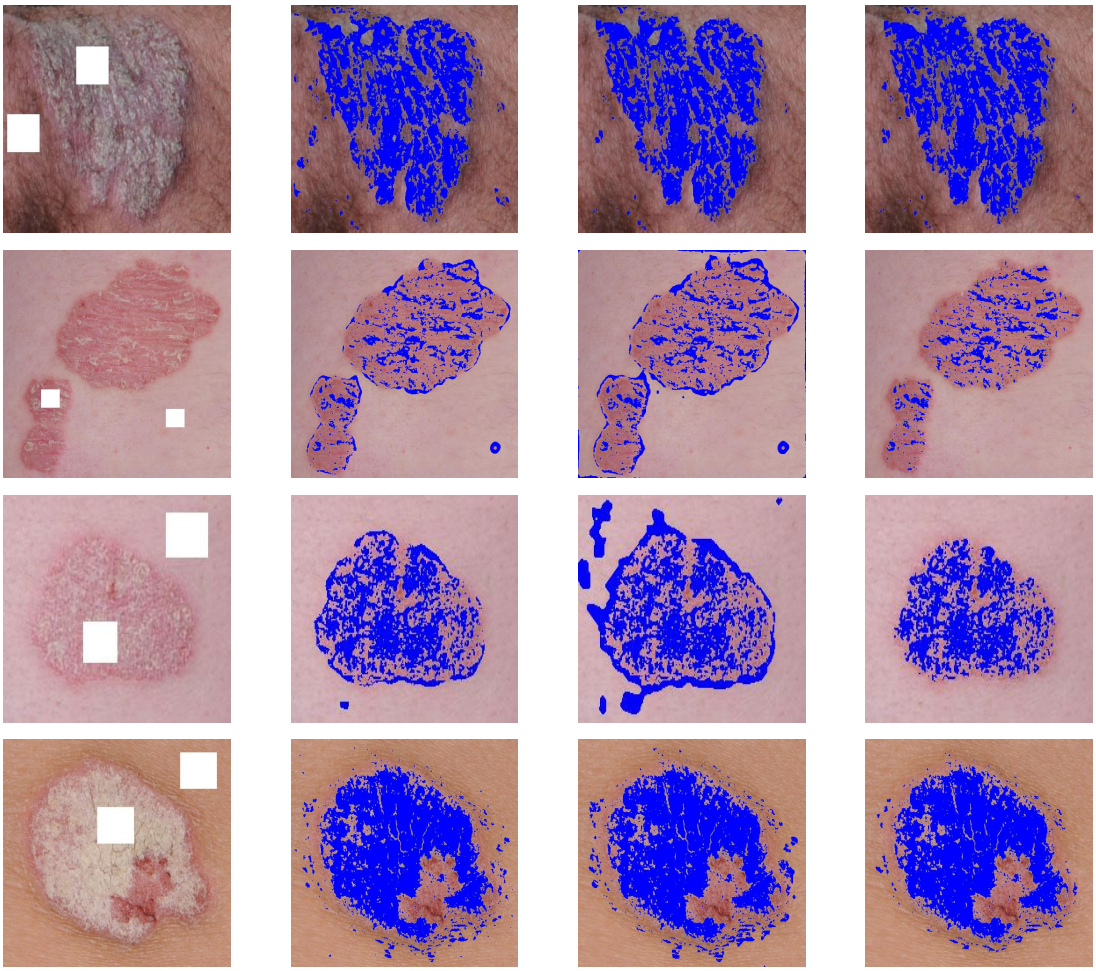


Figure 4.13: Classification results of the original images in Figure 4.11 with manually selected training sets. Detected scaling is marked in blue. (a) Selected training sets marked with white; (b) SVM segmentation; (c) MRF segmentation; (d) Our segmentation.

segmentation problem is converted to a binary classification problem by thresholding out erythema and other dark pixels. The training sets are automatically collected from individual images with the soft-constrained k-means. Thirdly, based on the SVM classification, the MRF is used to correct the misclassified pixels around the lesion borders by smoothing the SVM classification results. Finally, an experimental analysis is presented. The proposed algorithm is tested on psoriasis images with various skin colours, with or without wrinkles and hair, and with changes of illuminance conditions and photographing angles. The parameters, i.e. the threshold value for erythema removal, the probability

used in the soft-constrained k-means clustering, and the penalty constant k used in the MRF, are validated as well.

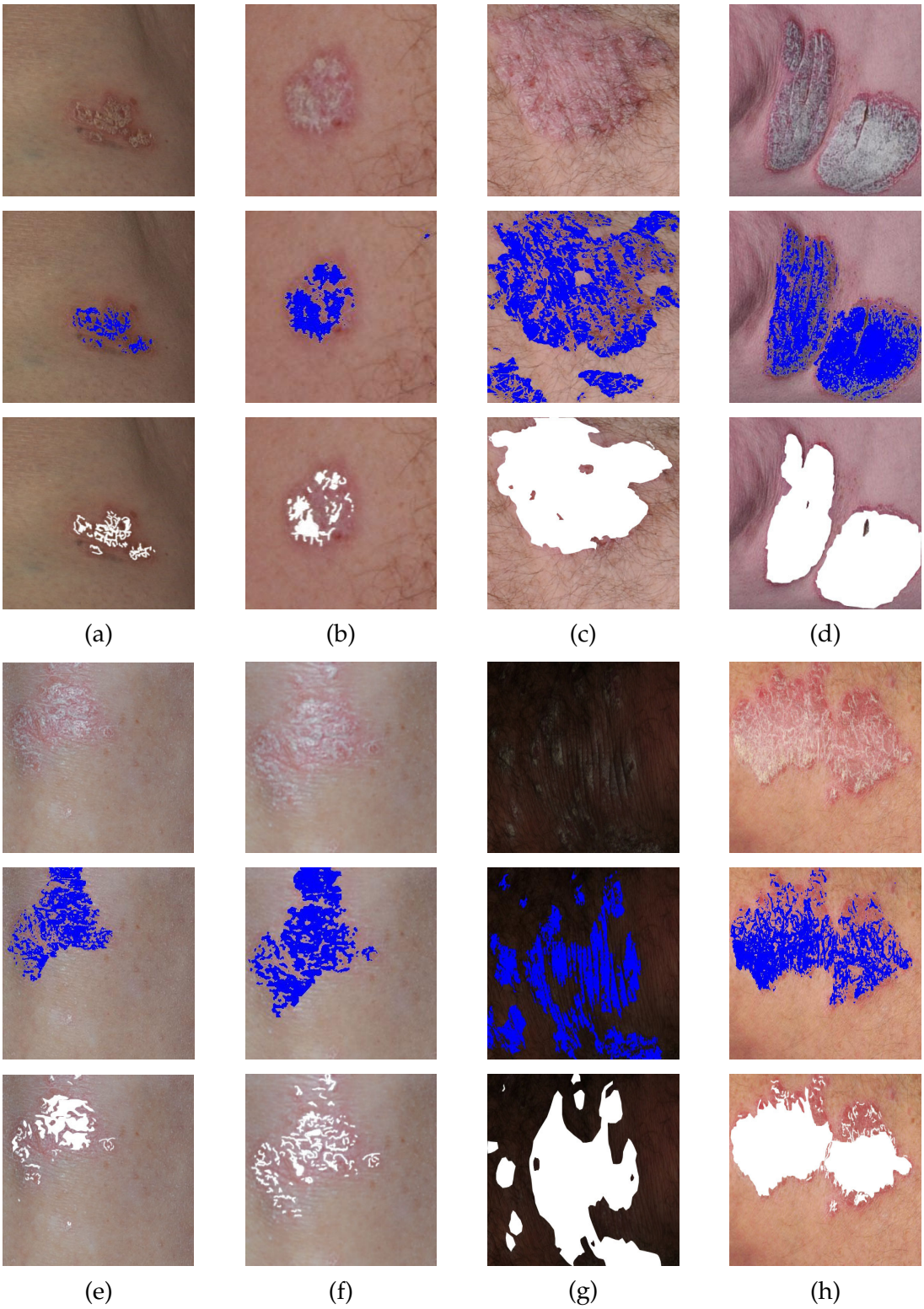


Figure 4.14: Segmentation results for a variety of scaling images. The first row in each group is the original image; the second row in each group is our segmentation result; the third row in each group is the ground truth. (a) Image with shadow; (b) Image with short hair; (c) Image with long hair; (d) Image with wrinkled skin; (e) Image captured from a certain angle; (f) Image captured with a different angle from the image in (e); (g) Image with a low illumination; (h) Image with a high illuminance.

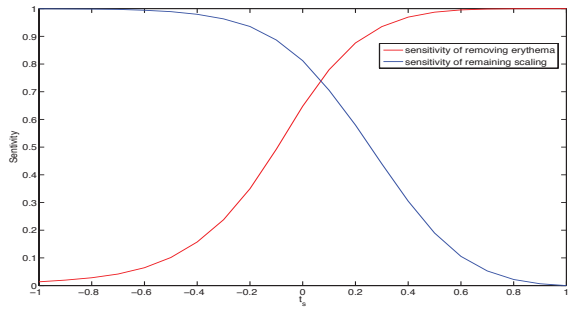


Figure 4.15: Sensitivity analysis of removing erythema and remaining scaling for variation of the threshold value t_S.

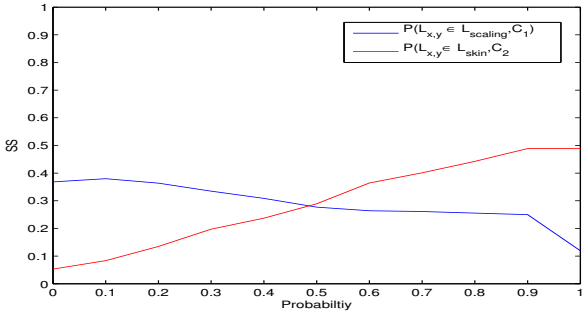


Figure 4.16: SS analysis of variation of the probability in the soft-constrained k-means.

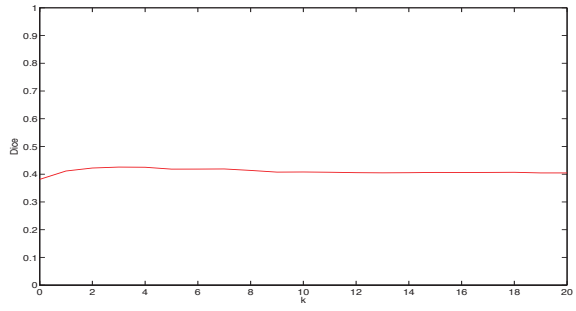


Figure 4.17: Dice analysis of variation of the penalty constant k in the SVM-based MRF classification.

Part IV

Treatment efficacy evaluation

Chapter 5

Erythema and Scaling Severity Assessment

5.1 Introduction

As introduced in Chapter 2, erythema and scaling are common symptoms in psoriasis. The severity of erythema and scaling is assessed in all the severity scoring systems [21–23], and all of these severity scoring systems use visual inspection to assess psoriasis severity, leading to unavoidable intra- and inter- observer variances. In this chapter, a computer aided-diagnosis system is proposed to reliably assess the severity of erythema and scaling by building on the erythema segmentation algorithm proposed in Chapter 3 and the scaling segmentation algorithm proposed in Chapter 4. In Figure 5.1, the flow chart of the proposed algorithm for psoriasis severity scoring is shown, where there are two key steps in this algorithm. The first one is psoriasis segmentation, that is composed of segmenting erythema and segmenting scaling; the second one is psoriasis severity scoring, where the severity of erythema and scaling is scored. The severity scoring algorithm depends on first being able to identify regions of psoriasis lesions.

The erythema and scaling severity given by a PASI scoring system is used as a benchmark to assess the performance of the computer-aided diagnosis system, considering that the PASI scoring system is the most popular severity assessment method in clinical research. In Table 5.1, the erythema severity scoring rule and scaling severity scoring rule in a PASI scoring system are listed. And the examples of psoriatic lesions with different severity levels are given. It can be noticed from Table 5.1 that the erythema severity and scaling severity levels are scored based on the appearances of the key symptom charac-

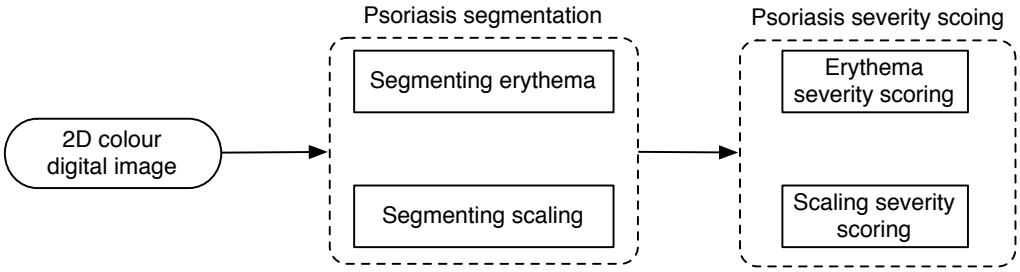
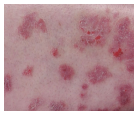


Figure 5.1: The key steps in the algorithm for psoriasis severity scoring.

ters of erythema and scaling.

This chapter begins with a presentation of the segmentation result of the whole psoriatic lesion using the erythema segmentation method described in Chapter 3 and the scaling segmentation method described in Chapter 4, followed by an introduction of the proposed computer-aided method for severity scoring. Next, the performance of the proposed method is analysed and validated through a linear correlation analysis and cross validation.

Table 5.1: PASI erythema and scaling severity intensity scoring.

Score	Grade	Erythema scoring	Erythema image	Scaling scoring	Scaling image
1	Mild	Light red		Fine scaling covering part of the lesion	
2	Moderate	Red,but not dark red		Fine to rough scaling covering a large part of the lesion	
3	Severe	Dark red		Rough, thick scaling covering a large part of the lesion	
4	Very severe	Very dark red		Very rough, very thick scaling totally covering the lesion	

5.2 Segmenting the whole psoriatic lesion in 2D digital skin images

Segmentation of the whole psoriatic lesion is a prerequisite for severity scoring of psoriasis. In this thesis, the segmentation is performed by segmenting erythema and scaling separately. In the rest of this section, we review the methods we proposed in Chapters 3 and 4 for the segmentation of erythema and scaling respectively.

In Chapter 3 there are two key steps in the segmentation of erythema: (**1**) the decomposition of skin colour $\mathbf{L}_{x,y}$ into the melanin component quantity $q^m_{x,y}$ and the haemoglobin component quantity $q^h_{x,y}$; and (**2**) the use of a Support Vector Machine (SVM) to separate erythema pixels from normal skin using the feature set composed of values of the haemoglobin and melanin quantities in an RGB colour space.

In Chapter 4, scaling is segmented from normal skin and erythema using a classification algorithm composed of the following steps:

Step 1 A scaling colour contrast filter is constructed using the L^* component and the a^* component in the $L^*a^*b^*$ colour space. The filter heightens the contrast between the white scaling pixels and the surrounding red erythema pixels. However, this filter does not always give enough contrast to differentiate scaling from normal skin, especially under bright lights or if the skin is pale like scaly skin in a lesion.

Step 2 The second step is to derive a set of Gabor responses for each pixel in the image using a bank of Gabor filters tuned for different directions and spatial frequencies. The resulting Gabor textures in the image can be displayed using a gray-scale value that captures the degree of '*roughness*' at the pixel. Scaling, which is rough, can be differentiated from normal skin, which is smooth, based on the Gabor textures.

Step 3 The third step is to use the Gabor textures together with the colour contrast features obtained from the scaling colour contrast filter to segment scaling from normal skin. The segmentation is performed by using an SVM smoothed by a Markov Random Field (MRF), which properly classifies pixels that are misclassified by the SVM.

Figure 5.2 shows the process of the segmentation of erythema and scaling of a psoriasis lesion from skin. It is observed that the whole psoriatic lesion is segmented out after identification of the erythema and the scaling regions.

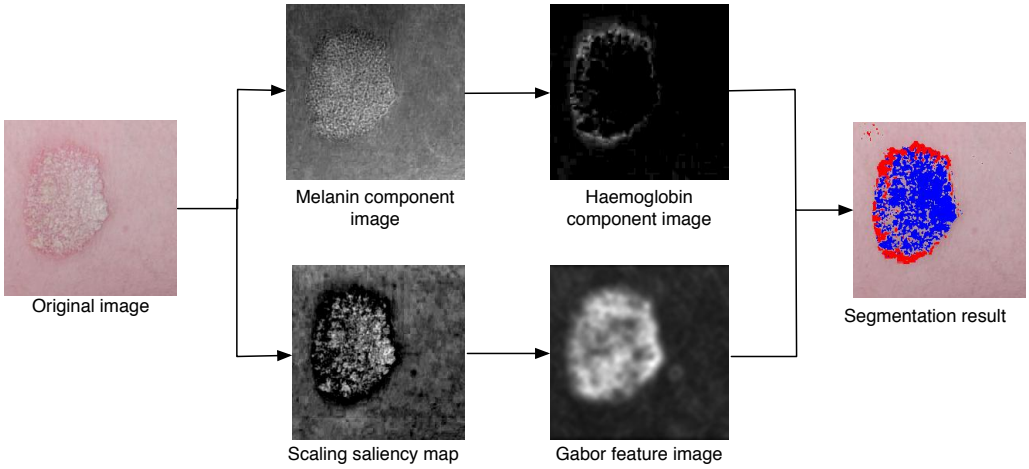


Figure 5.2: Segmentation of erythema and scaling, where segmented erythema is marked in red, and segmented scaling is marked in blue.

5.3 Scoring the severity of erythema and scaling

In this research, we view erythema and scaling scoring as a multiclass classification problem. The input is a 2D digital image of a lesion described by a set F_{severity} of severity features that depend on the symptom type that is being scored. Lesions with the severity score zero are not considered in this research, since there is no ambiguity between diagnosing skin with psoriasis and without psoriasis. The output of the multiclass classifier is a severity score in the set C_{severity}:

$$C_{\text{severity}} = \{1 := \text{Mild}, 2 := \text{Moderate}, 3 := \text{Severe}, 4 := \text{Very severe}\} \tag{5.1}$$

The classification function f_{severity} is developed to project a severity feature into a severity degree:

$$f_{\text{severity}} : F_{\text{severity}} \rightarrow C_{\text{severity}} \tag{5.2}$$

In this work, we use K-Nearest Neighbours (KNN) and C4.5 decision tree classifiers for the severity degree classification of erythema and scaling respectively. In the following of this section, we describe the two multiclass classifiers and their respective uses to model the severity scoring.

5.3.1 Multiclass classifiers: KNN and C4.5

K-nearest neighbours classification

The KNN method performs the classification by finding the k nearest training samples to the test sample. The class to which the test sample belongs is the one that includes most samples in the set of the k nearest training samples. Usually the Euclidean distance is used to find the k nearest training samples.

The KNN method has been widely used in biomedical engineering, especially for the severity scoring of a disorder that can be detected through imaging [62–65]. For example, in [62], a multifeature-based KNN method is applied to detect pulmonary embolisms on CT images; in [63] hormone receptors status in breast tissue sections are graded through KNN; in [64, 65] KNN methods are used for coronary calcification detection and scoring.

C4.5 decision tree classification

A decision tree partitions the feature space into different regions that are associated with different classes by using a series of classification queries. The classification queries make up nodes of the decision tree, and an answer of a query leads to a follow-up query that is the child node in the tree. Leaves of the tree are the classification results. This query-based modelling method directly shows how the classification results are determined by the classification features. In this work, a C4.5 decision tree algorithm is used to build the decision tree. Besides C4.5, there are many other decision tree methods, such as CART and ID3. The C4.5 outperforms CART and ID3 due to the high efficiency and low space complexity [155].

C4.5 builds a decision tree by two steps. In the first step, an initial decision tree is built by searching for optimal classification queries, that are called splitting rules, at each node.

In the second step, the decision tree is pruned to reduce over fitting of the classification model.

A C4.5 method identifies the optimal splitting rule by using the gain ratio that describes the effectivity of the splitting. An optimal splitting is identified, when the maximum gain ratio is achieved. The gain ratio is defined as:

$$\text{Gain ratio}(\mathcal{D}, \mathcal{T}) = \frac{\text{Gain}(\mathcal{D}, \mathcal{T})}{\text{Split Info}(\mathcal{D}, \mathcal{T})} \tag{5.3}$$

where $\text{Gain}(\mathcal{D}, \mathcal{T})$ is the information gained through a splitting $\mathcal{T}$. $\text{Gain}(\mathcal{D}, \mathcal{T})$ measures the change of impurity from a parent node $\mathcal{D}$ to its child nodes or leaves by using entropy. Its definition is given by:

$$\text{Gain}(\mathcal{D}, \mathcal{T}) = \text{Entropy}(\mathcal{D}) - \sum_{i=1}^{k} \frac{|\mathcal{D}_i|}{|\mathcal{D}|} \times \text{Entropy}(\mathcal{D}_i) \tag{5.4}$$

where $\mathcal{D}_i$ is a child node of $\mathcal{D}$ resulted from the splitting $\mathcal{T}$, k is the total number of subsets corresponding to the samples at node $\mathcal{D}$, and $|\cdot|$ is the number of samples at the corresponding note. When the gain is maximised, the impurity of the weighted subsets is lowest. The entropy gives a measurement of the impurity. If all samples at a node belong to the same class, the impurity of the subset is lowest, and so is the entropy. The definition of the entropy of a node $\mathcal{D}$ is given by:

$$\text{Entropy}(\mathcal{D}) = -\sum_{i=1}^{c} P(i|\mathcal{D}) \times \log_2 P(i|\mathcal{D}) \tag{5.5}$$

where $P(i|\mathcal{D})$ is the percentage of samples with class i at the node $\mathcal{D}$, and c is the total number of classes.

In Eq. (5.3), the $\text{Split Info}(\mathcal{D}, \mathcal{T})$ limits the number of subsets obtained by the splitting $\mathcal{T}$. Since when the number of subsets is big, the number of the associated sample in each subset are small, this kind of partition provides little information and is not preferred.

$$\text{Split Info}(\mathcal{D}, \mathcal{T}) = -\sum_{i=1}^{k} \frac{|\mathcal{D}_i|}{|\mathcal{D}|} \times \log_2(\frac{|\mathcal{D}_i|}{|\mathcal{D}|}) \tag{5.6}$$

Pseudocode of building the initial decision tree is presented in Algorithm 2. The algo-

rithm is illustrated based on Quinlan's book describing the C4.5 decision tree [155]. The gain ratio method is recursively used for the partitioned subsets to decide the splitting rules of the subsets. When all samples at a node satisfy a stopping condition, the node is changed into a leaf, and the samples at the node are assigned to the same class.

Algorithm 2 C4.5: building an initial decision tree [155].

Input: Training samples composed of feature values and class labels

Output: An initial decision tree

1: **if** All the samples at node $\mathcal{D}$ satisfy a stopping condition **then**
2: $\mathcal{D}$ is a leaf
3: **else**
4: $\mathcal{D}$ is a parent node
5: find a splitting rule that has the maximum gain ratio and split the samples into subsets.
6: **for** each subset **do**
7: run the tree building algorithm described from line 1 to line 10
8: add the result as a child node to the parent node $\mathcal{D}$
9: **end for**
10: **end if**

The tree stops growing, when a stopping condition is achieved. In the C4.5 algorithm, two stopping conditions are defined. One is that samples at a node belong to the same class, and the other is for the situation where no information is gained by any possible splitting.

After a tree is initially built, a pruning is performed to reduce over fitting of the tree that is caused by the lack of representative samples. Leaves with small samples are removed. Subtrees are replaced with new nodes or frequently used main branches, if this replacement reduces the estimated error [155].

Comparison of KNN versus C4.5

The common feature of the two multiclass classifiers is that they are both non-parametric classifiers. In other words, no space distribution functions are required to estimate in the classification process.

Three main differences exist between KNN and C4.5:

1. One difference is that KNN is a black box classifier, while the classification of C4.5 is intuitive. When a KNN classification is performed, the detailed classification process is blind for a specific test sample. On the contrary, C4.5, showing classification rules, provides straightforward understanding of the classification process.
2. Another difference is the robustness of the two classification methods. C4.5 is less sensitive to noise than KNN, since the pruning technique is performed in the tree building stage.
3. Additionally, the two methods differ in computational complexity. The computational complexity of KNN is mainly due to the testing stage, and has the value of $O(mn)$, where m is the number of features and n is the number of training samples. The computational complexity of C4.5 has the value of $O(mn^2)$, that is mainly due to deducing classification rules in the training stage.

In practice, the choice of the two classifiers depends on the specific classification problems. We will use KNN for erythema scoring and C4.5 for scaling scoring, because these techniques give the best results for the two specific problems, as we will demonstrate in Section 5.4.

5.3.2 Erythema Severity modelling

The PASI guidelines for scoring erythema severity is based on the erythema colour (as shown in Table 5.1). The deeper the red colour, the higher the severity score. In this work, we use the haemoglobin and melanin quantities that we proposed in erythema segmentation (Chapter 3) to describe the erythema colour, and therefore the erythema severity.

Note that dermatologists score erythema by comparing with the surrounding normal skin. This is because people with different ethnic background have different skin colour, the colour of erythema with the same severity score may differ as well. The comparison with the surrounding normal skin can be considered as a standardisation, so that the erythema severity is measured against the same benchmark. Our algorithm uses a similar contrast technique by employing the relative haemoglobin and melanin quantity features, $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$, that are modelled by using a difference operation:

$$\begin{aligned} \nabla\overline{q^h} &= \overline{q_E^h} - \overline{q_S^h} \\ \nabla\overline{q^m} &= \overline{q_E^m} - \overline{q_S^m} \end{aligned} \tag{5.7}$$

where $\overline{q_E^h}$ and $\overline{q_S^h}$ are the mean value of the haemoglobin quantities of erythema and normal skin respectively, and $\overline{q_E^m}$ and $\overline{q_S^m}$ are the mean value of the melanin quantities of erythema and normal skin respectively.

The erythema severity feature set is given by:

$$F_{\text{severity}} = \{\nabla\overline{q^h}, \nabla\overline{q^m}\} \tag{5.8}$$

In this work, scoring of erythema is done by a KNN approach. We construct a feature space composed of the mean difference of haemoglobin values $\nabla\overline{q^h}$ and melanin values $\nabla\overline{q^m}$. The distance between a training sample and the sample to be classified is calculated using the Euclidean distance d_q:

$$d_q = \sqrt{(\nabla\overline{q^h}_i - \nabla\overline{q^h}_o)^2 + (\nabla\overline{q^m}_i - \nabla\overline{q^m}_o)^2} \tag{5.9}$$

where $\overline{q^h}_i$ and $\overline{q^m}_i$ are the erythema severity features of the ith sample in the training set, and $\overline{q^h}_o$ and $\overline{q^m}_o$ the erythema severity features of the test sample.

The k nearest training samples of the test sample are decided by using the Euclidean distance as defined in Eq. (5.9). In this work $k = 5$ has been empirically determined.

5.3.3 Scaling Severity Modelling

The PASI scoring of scaling is also shown in Table 5.1. It is observed that the severity score for scaling is based on the roughness of scaling and the area of scaling relative to the whole lesion. Thus, we use the Gabor texture, that is defined to differentiate rough scaling from smooth normal skin in Chapter 4, to describe the roughness of scaling, and the ratio of scaling area to the whole lesion to describe the relative scaling area.

The severity features used to differentiate scaling from erythema and normal skin are

modelled as:

$$F_{\text{severity}} = \{\overline{g}, r\} \tag{5.10}$$

where $\overline{g}$ is a roughness feature, that is the mean value of the Gabor textures, and r is the relative scaling area that is the ratio of scaling area to the whole lesion area.

To classify the scaling severity degree, a decision tree is built by using the C4.5 algorithm. The resulting decision tree is composed of a series of splitting rules, that use threshold values of the mean value of the Gabor textures $\overline{g}$ and the relative scaling area r to perform the classification.

5.4 Validating the Method

Samples of psoriasis skin images are collected from the Skin & Cancer Foundation Victoria, where the imaging environment is carefully set to ensure controlled illumination. The images used in the comparison were those that were given the same PASI scores by two dermatologists.

The proposed algorithm is validated by using linear correlation and classification performance analyses. In the linear correlation analysis part, an F-test is conducted to examine the difference of severity features between any two different severity groups. The linear correlation between the severity features with the severity scores given by dermatologists is analysed by using a correlation coefficient. In the classification performance analysis part, the classification performance is evaluated by comparing the true classification accuracy of our scoring method with other scoring systems using cross validation.

Two series of experiments were separately performed. The first evaluated the performance of erythema severity scoring method by using the linear correlation and classification performance analyses; the second conducted the same performance evaluation methods on our proposed scaling severity scoring method.

5.4.1 The Experimental Design

Experimental design of the linear correlation evaluation

In the linear correlation stage, an F-test in a one-way analysis of variance is used to test the difference between any two groups of severity features that are corresponding to different severity scores. The following hypotheses are examined:

$$\begin{aligned} \mathcal{H}_1: &\quad \text{at least one } i \neq j \quad \mu_i - \mu_j \neq 0 \\ \mathcal{H}_0: &\quad \text{for any } i \neq j \quad \mu_i - \mu_j = 0 \end{aligned} \tag{5.11}$$

where μ_i and μ_j are the means of severity features belonging to different groups of severity scores, we have $i, j \in \{1, 2, 3, 4\}$ for four degrees of severity, $\mathcal{H}_0$ is the null hypothesis, and $\mathcal{H}_1$ is the alternative hypothesis.

The resulting p-values and F-values of the F-test are analysed. The p-value is the probability of the F-test, when the null hypothesis is true. A low p-value indicates the means of two classes are different significantly. Generally, it is considered that a significant p-value should be less than 5%.

An F-value is used to measure the variance of difference between groups of two classes. It is the ratio of between-group variability to within-group variability.

$$\begin{aligned} \mathrm{F-value} &= \frac{\text{between-group variability}}{\text{within-group variability}} \\ &= \frac{\frac{\sum_{i=1}^{k} |f_i|(\bar{f}_i - \bar{f})^2}{k-1}}{\frac{\sum_{i=1}^{k}(|f_i|-1)\delta_i^2}{(\sum_{i=1}^{k}|f_i|)-k}} \end{aligned} \tag{5.12}$$

where $\bar{f}_i$ and δ_i are respectively the mean value and the standard deviation of the feature set with severity score i, and $\bar{f}$ is the mean of all samples. Since four kinds of severity scores are considered in the experiment, k is equal to 4, the degrees of freedom for the numerator is 3, and the degrees of freedom for the denominator is $(\sum_{i=1}^{k} |f_i|) - k$.

If an F-value is larger than one, larger the F-value is, higher distinguishability of the the two groups is. It indicates the between-group variability is much larger than the within-group variability. Thus, the null hypothesis is skeptical.

Though the F-test indicates that the severity features between different severity degrees are significantly different, we still need to know how the severity features extracted from the lesions with different severity scores are correlated with these severity scores. The linear correlation relationship between severity scores and severity features is analysed using a correlation coefficient, that is defined by using a normalised covariance:

$$\text{Correlation coefficient} = \frac{\text{cov}(f, s_m)}{\sqrt{\text{cov}(f, f) \times \text{cov}(s_m, s_m)}} \tag{5.13}$$

where $\text{cov}(\cdot)$ is a covariance operation, f is a variable of a severity feature, and s_m is a variable of the corresponding severity score. The correlation coefficient gives a dependency measurement to the tested feature set and the severity score set. A higher correlation coefficient indicates a higher dependency between f and s_m. If the value is close to 1, it implies that the severity feature f and the severity score s_m have a positive linear relationship. If the value is close to -1, an increase of one variable is accompanied by a decrease of the other.

The p-value of the correlation relationship is analysed. In this case, the null hypothesis is that the severity features are not correlated with the severity degrees. When the p-value is less than 0.05, the null hypothesis is rejected, and the correlation analysis is significant.

Experimental design of the classification performance evaluation

In the classification performance evaluation stage, the performance of a severity classifier is evaluated by using the classification accuracy, which is defined as the percentage of correctly classified psoriatic lesion samples. Additionally, in order to compare the performances of the classifiers without bias, 10-fold cross validation is applied. The true accuracy difference between two classifiers is analysed by using the classification results in the 10-fold cross validation.

In 10-fold cross validation, the classification performance is analysed by randomly dividing the collected sample set into 10 groups of equal-sized subsets. Each group is in turn used as a testing set, and the rest groups are used as training samples. The classification accuracy of each fold for different severity degrees is calculated.

The true accuracy difference is about the statistical significance of the accuracy difference observed during the 10-fold cross validation. It gives an estimation of the true accuracy difference between two classifiers under a certain confidence level. If zero is not in the range of the true accuracy difference, the performances of the two classifiers are significantly different. In this work, assuming the accuracy difference between two classifiers is a t-distribution, by setting the confidence level to be 0.95, the confidence interval of true accuracy difference $\tilde{d}$ is given by:

$$\tilde{d} = \bar{d} \pm t_{(0.95),10-1}\delta_d \tag{5.14}$$

where $\bar{d}$ is the mean accuracy difference between the two classifiers across the 10-fold validation, $t_{(0.95),10-1}$ is the probability of t-distribution with confidence level 0.95 and the degrees of freedom $10-1$, and δ_d is the overall variance of the difference that is estimated by:

$$\delta_d = \frac{\sum_{j=1}^{10}(d_j - \bar{d})^2}{10(10-1)} \tag{5.15}$$

where d_j is the accuracy difference of the two classifiers at the jth fold.

5.4.2 Erythema Severity Scoring

There are 88 images of erythema lesions spanning a number of different skin types and ethnicities, where 10 samples score 1, 31 samples score 2, 37 samples score 3, and 10 samples score 4.

Linear Correlation Analysis of erythema severity features

The result of linear correlation between erythema severity and colour features is shown in Table 5.2. Table 5.2 also shows relative hue component $\nabla\overline{H^{ab}}$ used in [32], relative colour component $\nabla\overline{a^*}$ derived from the L*a*b* colour space proposed in [101], and relative colour components $\nabla\overline{R}$, $\nabla\overline{G}$, $\nabla\overline{B}$ from the RGB colour space proposed in [7]. The relative colour component features are defined to be the difference of mean values between erythema and normal skin. Specifically, the hue component in [32] is explained in the L*a*b* space. It is defined as the arctangent of ratio of the $\mathbf{b}^*$ component to the $\mathbf{a}^*$ component:

$$h_{a^*b^*} = \arctan(\frac{\mathbf{b}^*}{\mathbf{a}^*}) \tag{5.16}$$

The distribution of the relative melanin and haemoglobin features with changes to erythema severity intensity is presented with box plots in Figure 5.3. Notice that though mean values of the relative melanin feature, relative haemoglobin feature and their summation are linearly related with the severity intensity in general, it is hard to distinguish erythema severity by setting thresholding values alone, since the value ranges of the two features and their summation are overlapped.

Table 5.2: Analysis of linear correlation between erythema severity and colour features.

Features	p-value of the F test	F-value	p-value of the correlation analysis	Correlation coefficient
$\nabla\overline{H^{ab}}$	0.299	1.24	0.1154	0.048
$\nabla\overline{a^*}$	2.20E-07	13.77	0	0.57
$\nabla\overline{R}$	4.52E-06	10.79	0	-0.448
$\nabla\overline{G}$	2.18E-10	21.4	0	-0.63
$\nabla\overline{B}$	6.19E-07	12.72	0.0851	-0.538
$\nabla\overline{R} + \nabla\overline{G} + \nabla\overline{B}$	2.10E-08	16.22	0.1186	-0.569
$\nabla\overline{q^m}$	0.048	2.74	0.0732	0.194
$\nabla\overline{q^h}$	1.44E-09	19.19	0	0.616
$\nabla\overline{q^h} + \nabla\overline{q^m}$	6.58E-06	10.439	0	0.456

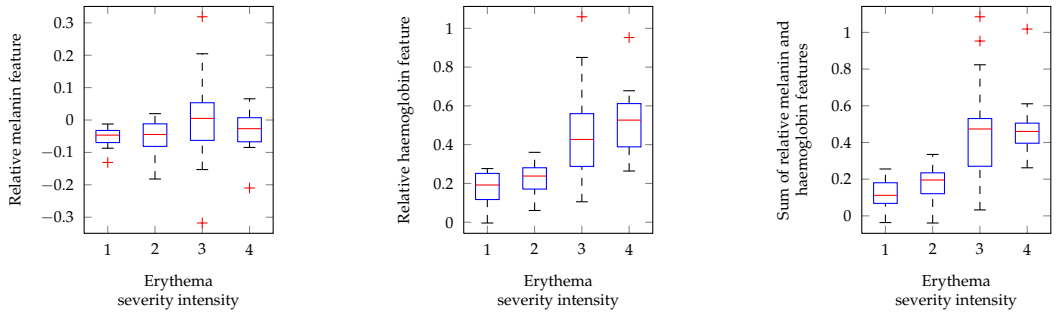


Figure 5.3: Box plots of distribution of relative haemoglobin and melanin features with erythema severity score. (a) Box plot of distribution of relative melanin feature with erythema severity score. (b) Box plot of distribution of relative haemoglobin feature with erythema severity score. (c) Box plot of distribution of summation of relative haemoglobin and melanin features with of erythema severity score.

Classification performance analysis of erythema severity scoring

The result of our erythema scoring method is compared with the result obtained by the C4.5 decision tree algorithm [155] and the Minimum Centre Distance (MCD) algorithm [40], where a test sample is grouped into a class in which the mean value of the training samples is closest to the test sample in a feature space.

Table 5.3 illustrates the classification accuracy of KNN, C4.5 and MCD classification methods by using the features of relative colour component $\nabla\overline{R}$,$\nabla\overline{G}$ and $\nabla\overline{B}$ in an RGB colour space and the features of relative haemoglobin and melanin components, $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$. And, Figure 5.4 shows the accuracy for different severity scores in the 10-fold cross validation analysis. A KNN classification method using the relative haemoglobin and melanin components shows the highest accuracy compared with other methods. The mean accuracy in 10-fold cross validation is 77.71%. Moreover, the performance of KNN is better than MCD and C4.5 in general. For erythema, the accuracy of severity 1 and severity 4 are better than the accuracy for severity 2 and severity 3. This is due to the ambiguity in scoring severity 2 and severity 3 lesions.

Assuming that the accuracy differences in the 10-fold cross validation analysis is a t-distribution, the statistical significance of the accuracy difference between the top classifier using KNN with the features used in our algorithm and other severity classifiers is shown in Table 5.4. The indications are that performance of the top classifier is signifi-

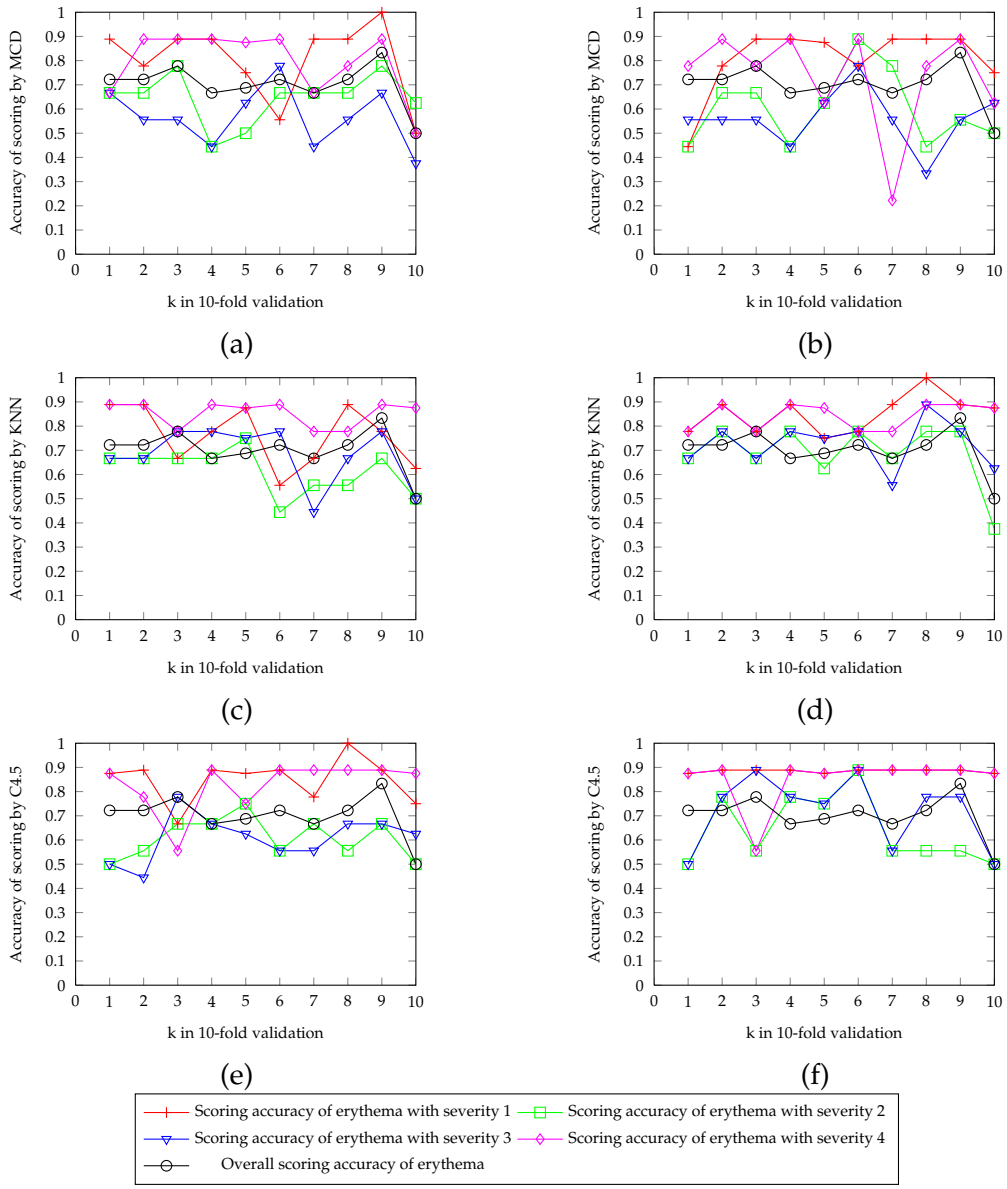


Figure 5.4: Accuracy of erythema scoring with 10-fold cross validation for a series of severity and the corresponding classifier. (a) Accuracy of erythema scoring using MCD and the severity features: $\nabla\overline{R}$, $\nabla\overline{G}$, and $\nabla\overline{B}$. (b) Accuracy of erythema scoring using MCD and the severity features: $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$. (c) Accuracy of erythema scoring using KNN and the severity features:$\nabla\overline{R}$, $\nabla\overline{G}$, and $\nabla\overline{B}$. (d) Accuracy of erythema scoring using KNN and the severity features: $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$. (e) Accuracy of erythema scoring using C4.5 and the severity features:$\nabla\overline{R}$, $\nabla\overline{G}$, and $\nabla\overline{B}$. (f) Accuracy of erythema scoring using C4.5 and the severity features: $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$.

cantly better than the rest.

Table 5.3: Classification accuracy of erythema severity scoring.

Severity feature	Method		
	MCD	KNN	C4.5
$\nabla\overline{R}, \nabla\overline{G}, \nabla\overline{B}$	0.7021	0.7271	0.7236
$\nabla\overline{q^h}, \nabla\overline{q^m}$	0.6757	0.7771	0.7743

Table 5.4: Accuracy difference between the classifier using KNN with the severity features of relative haemoglobin and melanin features, $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$, and other implemented classifiers.

Compared classifier	Colour space	Accuracy difference
MCD	RGB colour space	[0.0738, 0.0762]
MCD	Haemoglobin and melanin component	[0.098, 0.103]
KNN	RGB colour space	[0.049, 0.051]
C4.5	RGB colour space	[0.0053 0.0054]
C4.5	Haemoglobin and melanin component	[0.0021 0.0034]

5.4.3 Scaling Severity Scoring

In the scaling severity scoring part, 52 images of psoriatic lesion are collected. They are composed of 10 images with scaling severity 1, 17 images with scaling severity 2, 18 images with scaling severity 3, and 7 images with scaling severity 4.

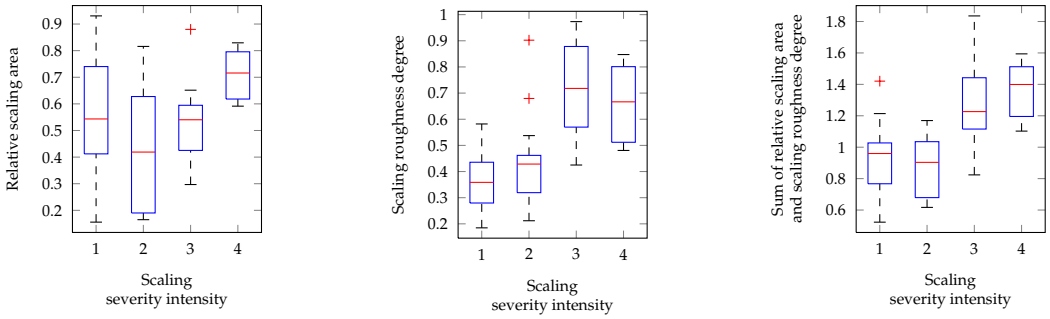
Linear correlation analysis of scaling severity features

The result of linear correlation analysis of scaling severity features is shown in Table 5.5. We examined the feature scaling area s_a, which is considered as a major element related with scaling severity in [7], and the feature set proposed in this work.

Figure 5.5 shows the distribution of relative scaling area, scaling roughness degree and their summation with changes of scaling severity intensity. We can see that ranges of the two features and their summation overlap for different scaling severity intensities. No clear linear relationship exists between the features and the severity intensities.

Table 5.5: Analysis of linear correlation between scaling severity and its features.

Features	p-value of the F test	F-value	p-value of the correlation analysis	Correlation coefficient
s_a	0.023	3.44	0.0157	0.316
r	0.019	3.61	0.1772	0.18
$\overline{g}$	8.67E-09	19.81	0	0.641
r+$\overline{g}$	4.59E-07	14.55	0	0.58



(a) (b) (c)

Figure 5.5: Box plots of relative scaling area and scaling roughness degree distribution with changes of scaling severity intensity. (a) Box plot of relative scaling area distribution with changes of scaling severity intensity. (b) Box plot of scaling roughness degree distribution with changes of scaling severity intensity. (c) Box plot of summation of relative scaling area and scaling roughness degree distribution with changes of scaling severity intensity.

Classification performance analysis of scaling severity scoring

The performance of our scaling scoring method is compared with a KNN method and the method in [7], where a decision tree is employed to score the severity degree with the scaling area. The classification accuracy result is shown in Table 5.6. In Figure 5.6, the accuracy with 10-fold cross validation for difference severity scores is illustrated. For KNN and the decision tree, scoring accuracy for severity 1 and severity 4 is generally better than scoring accuracy for severity 2 and severity 3 as with erythema. Moreover, using the feature set of our method the accuracy is much better than features proposed in [7].

In Table 5.7, the statistical significance of the accuracy difference between the classifier using the decision tree with the relative scaling area and the scaling roughness degree and other implemented scaling scoring classifiers is shown. All of the difference ranges

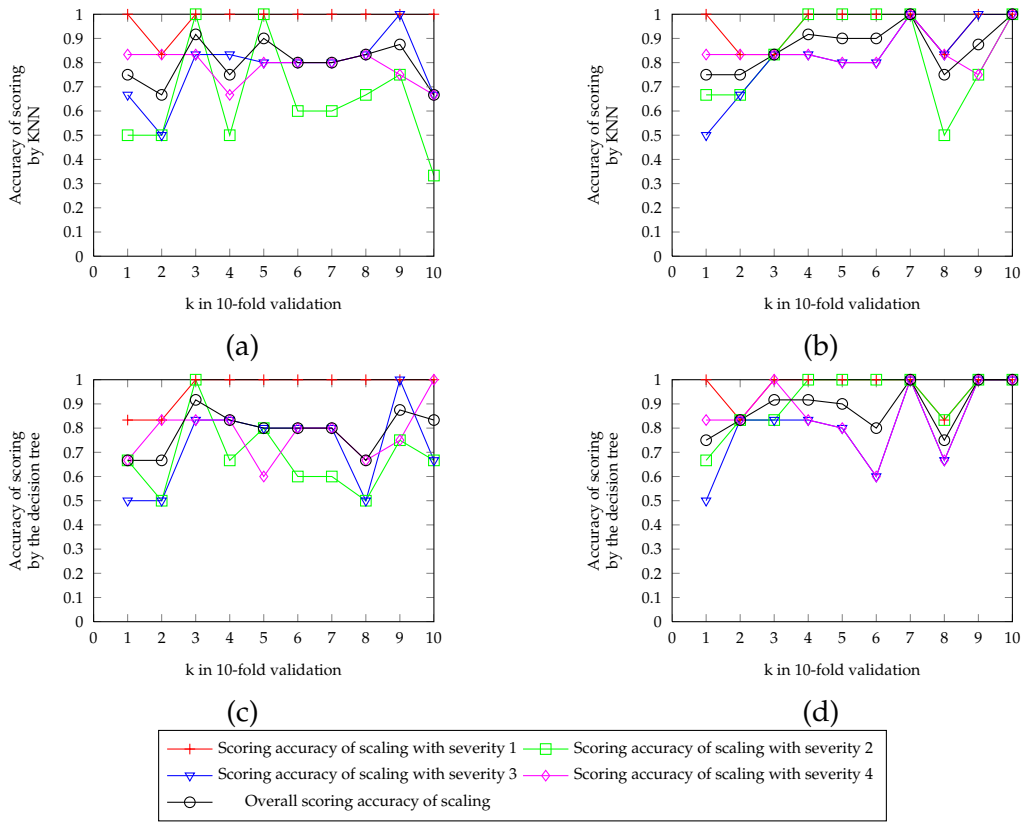


Figure 5.6: Accuracy of scaling scoring with 10-fold cross validation for a series of severity and the corresponding classifier. (a) Accuracy of scaling scoring using KNN with the feature proposed in [7]. (b) Accuracy of scaling scoring using KNN with the features of relative scaling area and scaling roughness degree. (c) Accuracy of scaling scoring using the decision tree with the feature proposed in [7]. (d) Accuracy of scaling scoring using the decision tree with the features of relative scaling area and scaling roughness degree.

do not include zero. Thus the accuracy of the decision tree we proposed is significantly different when compared with the other classifiers.

5.4.4 Discussion of experiments

We experimented our methods on psoriatic lesions with erythema severity and scaling severity ranging from the lowest to the highest, except for non-symptoms. The proposed erythema and scaling severity features correlate well with dermatologists' diagnoses. Both erythema scoring and scaling scoring give good classification accuracy for the lowest and highest severity, while the classification accuracy for medium severity is a little lower.

Table 5.6: Classification accuracy of scaling severity scoring.

Severity feature	Method	
	KNN	C4.5
s_a	0.7958	0.7858
$r, \overline{g}$	0.8675	0.8867

Table 5.7: Accuracy difference between the classifier using the decision tree with relative scaling area and scaling roughness degree and other implemented scaling scoring classifiers.

Compared classifier	Scaling features	Accuracy difference
KNN	Features in [7]	[0.087, 0.095]
KNN	Relative scaling area and scaling roughness degree	[0.081, 0.02]
Decision tree	Features in [7]	[0.1, 0.102]

For the erythema severity scoring, by comparing with previously proposed erythema scoring models, the severity features proposed in our method is much better, especially than the severity feature proposed in [32]. As indicated In Table 5.2, statistical analysis values of our method are much higher.

In the erythema severity classification part, in Figure 5.4 and Table 5.3, it is shown that the scoring performance of using relative melanin and haemoglobin components is better than using relative colour components from the RGB colour space [7]. This is attributed to that the haemoglobin and melanin components derived from the ICA decomposition are directly related with skin colouring. Though melanin is not about redness of skin, it is still considered in erythema severity scoring, because normal skin colour decided by melanin is referred to obtain erythema severity intensity. Additionally, the performance of erythema scoring by using KNN outperforms MCD and C4.5. This is because the linear classifier MCD is not suitable to eliminate classification bias caused by the overlapping of feature values of different severity, and C4.5 suffers the same deficiency. Though C4.5 is a non-linear classifier, since C4.5 divides the feature space into several regions whose edges are parallel to the axises, this classifier is not good at dealing with the feature values that are not very well separated for different severity scores. However, KNN is superior, since this method does not use strict threshold values. KNN adaptively decides the severity

degree of a test sample by considering the majority of severity degrees of training samples in its neighbourhood. This method is good at correcting the misclassification of test samples, whose feature valuses are in the overlapped area of feature values from different classes. In this case, the test sample is more likely to be misclassified by MCD and C4.5. For KNN, when the majority training samples in the neighbourhood is not in the overlapped area or, even though the traning samples in the neighbourhood in the overlapped area, the majority tranining samples come from the same class as the test sample, KNN is able to correctly decide the seveirty degree of the test sample.

For the scaling severity scoring, our scaling severity scoring method has the best performance among the tested methods. The comparison of scaling severity features is shown in Table 5.5. It is observed that even though scaling area has a higher correlation coefficient than relative scaling area, it does not possess priority in p-values and F-values, especially when compared to the roughness degree.

In the scaling severity classification part, our scaling severity scoring method outperformed the method in [7], where only scaling area is used. Table 5.5 illustrates the reason by showing that the p-values and F-values of our proposed severity features are better than the feature of scaling area used in [7]. Additionally, by using the proposed severity features, the performance of C4.5 is better than KNN. Though C4.5 has a lower performance in the experiment of erythema severity scoring, C4.5 has a better scaling severity scoring result, since the proposed scaling severity features are more highly correlated with the corresponding severity scores. This can be examined by comparing the linear correlation analysis result of the proposed scaling severity features in Table 5.5 with the result of the proposed erythema severity features in Table 5.2. The low performance of KNN is mainly due to samples with feature values close to the decision boundary. When there are more samples belonging to another severity degree around the decision boundary, the severity degree of a test sample is misclassified. Since the correlation coefficents of the scaling severity features is better than the erythema severity features, it means that there are less scaling severity feature values than erythema severity feature values in the overlapped area of feature values from different classes. In this case, KNN does not out outperform C4.5.

In this work, we propose an automatic method to assess the severity of erythema and scaling, that avoids the disadvantages of traditional manual severity evaluation method. The accuracy of the proposed severity assessment methods is above the accepted 75%. This performance satisfies the requirement of a computer aided diagnosis system. However, in future, other colour features, that shows good correlation with erythema severity degrees, can be exploited to improve the performance of erythema seveirty assessment, e.g. the **G** value in the RGB colour space. Texture analyasis methods, such as Laplacian of Gaussian that measures colour intensity changes using Laplacian and Gaussian filters together, can be researched to improve the correlation between the severity degrees and the severity features of scaling. Moreover, it is possible to improve the accuracy of severity classification by using more advanced machine learning methods, e.g. SVMs and artifical neural networks.

5.5 Summary

This chapter presents an algorithm of using severity features extracted from psoriasis skin images to reliably evaluate the psoriasis severity. The psoriasis severity is evaluated from the aspects of erythema severity and scaling severity. The erythema severity is described by the contrasts of haemoglobin and melanin values in psoriatic lesions to the corresponding values in the surrounding skin, and scored by using KNN. The scaling severity is described by the relative scaling area and the scaling roughness degree modelled with the Gabor texture feature. The C4.5 decision tree is used to score the scaling severity. The superiority and reliability of the severity features and the scoring methods are validated by comparing with previous severity models.

Chapter 6

Severity Changes Assessment and Its Clinic Practice

In this chapter, the efficacy of psoriasis treatment is assessed from the aspect of the changes in severity of a lesion during a treatment period. The contribution of this chapter is to develop a method to assess the severity changes without registration of psoriatic lesion images. As noted in Chapter 2, previous work on the assessment of changes relies on the registration of lesions in different images, which is only available for short-term treatments, since boundaries of psoriatic lesions are almost the same in before-after psoriasis images photographed in short-term treatments. In long-term treatments, not only dose the content of the lesion change, but also boundaries of the same lesion can dramatically change. Therefore, it is hard to conduct lesion registration in this situation. In Figure 6.1, an example of two images of the same lesion taken one month apart is shown.

Since in Chapter 5, we shows that our proposed erythema and scaling severity features have a good linear correlation with the corresponding severity scores, we model

(a) (b)

Figure 6.1: An example of the before-after images taken in a psoriasis treatment. (a) The before image, (b) The after image photographed after one month.

the severity changes here by using a linear regression analysis factoring in the changes of severity features proposed in Chapter 5. By doing this, the severity changes are assessed without image registration.

This chapter begins with a presentation of this method, including descriptions of the proposed severity change features and modelling with multiple linear regression analysis. This is followed by an analysis of the consistency between the severity change models and severity scores evaluated by clinicians in the experiments. It is shown that using the proposed method to assess the severity changes is reliable.

6.1 Severity Change Modelling

In this section severity change features for erythema and scaling are extracted first. Next, building on the extraction of severity change features, we use a multiple linear regression analysis to model the severity changes evaluated by using PASI. The linear regression analysis provides a way to examine the linear relationship between the severity change features and the changes of severity scores given by PASI.

6.1.1 Severity Change Features of Erythema and Scaling

The severity change features are constructed using the severity features presented in Chapter 5. There are two kinds of severity change features. One is the erythema severity change features derived from the haemoglobin and melanin components of the lesion; the other is the scaling severity change features associated with Gabor features and the relative scaling area.

Changes in lesion severity are described by a subtraction between the severity features of a lesion at one time point and the features of the same lesion at another point in time. A general severity change feature is constructed from a function $D(X)$ given by:

$$D(X) = X_{\tau 2} - X_{\tau 1} \tag{6.1}$$

where X_{τ_2} is the severity feature at the second time point, and X_{τ_1} is the severity feature

at the first time point.

The change in erythema severity within a lesion can now be defined by the **erythema severity change feature set** C_{Erythema}, which is related with changes of relative haemoglobin quantities $D(\nabla\overline{q^h})$ and changes of relative melanin quantities $D(\nabla\overline{q^m})$ (see Section 5.3.2 for the details of relative haemoglobin and melanin quantity features):

$$C_{\text{Erythema}} = \{D(\nabla\overline{q^h}),\ D(\nabla\overline{q^m})\} \tag{6.2}$$

Since the scaling severity depends on two factors: the roughness of the scaling and the area of scaling relative to the whole lesion, the changes in scaling severity are modelled by the **scaling severity change feature set** C_{Scaling} that is composed of the changes in the degree of roughness $D(\overline{g})$ and relative areas $D(r)$ (see Section 5.3.3 for the details of) :

$$C_{\text{Scaling}} = \{D(r),\ D(\overline{g})\} \tag{6.3}$$

6.1.2 Modelling with multiple linear regression analysis

In order to find how the severity change features for erythema and scaling are related with the changes of corresponding severity scores in PASI, a multiple linear regression analysis is used in this work. The multiple linear regression model assumes that the relationship between the severity change features and the severity change degrees is linear. A linear function parameterised with the severity change features is estimated in the regression analysis.

A general multiple linear regression model is expressed as:

$$\hat{y}_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \alpha \tag{6.4}$$

where $\hat{y}_i$ is the ith estimated change of severity score corresponding to the two ith elements in the severity change feature set $x_{1,i}$ and $x_{2,i}$, β_1 and β_2 are the coefficients associated with the severity change features, and α is the intercept. The coefficients β_1 and β_2, and the intercept α are found by minimising least square differences between the estimated changes of severity score and the true changes of severity score. We have:

$$[\beta_1, \beta_2, \alpha]^t = (([\mathbf{x}_1, \mathbf{x}_2, I_{n\times 1}]^t[\mathbf{x}_1, \mathbf{x}_2, I_{n\times 1}])^{-1}[\mathbf{x}_1, \mathbf{x}_2, I_{n\times 1}]^t\mathbf{y}) \tag{6.5}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors with n observed severity feature samples: $\mathbf{x}_1 = [x_{1,1}, x_{1,2}, ..., x_{1,n}]^t$ and $\mathbf{x}_2 = [x_{2,1}, x_{2,2}, ..., x_{2,n}]^t$, $I_{n\times 1}$ is an n element vector with n 1s, and $\mathbf{y} = [y_1, y_2, ..., y_n]^t$.

6.2 Experimental Validation

Psoriasis skin images are collected from the Skin & Cancer foundation Victoria, where the imaging environment is carefully set to ensure controlled illumination. The set of images chosen include various skin types from Asian and Caucasian ethic backgrounds.

The images for a specific lesion were collected at two different time points and given PASI scores by two dermatologists. Only those images that were given identical PASI scores by the two dermatologists for both time points were selected. For comparison with the algorithms, the ground truth is chosen to be the difference in PASI severity scores between the two time points for each lesion. We note that a straight subtraction of severity scores between two time points may yield a negative value indicating a decrease in severity, or a positive value indicating an increase in severity. The situation is symmetrical for our analysis. In this case, only severity decrease is considered in the experiment. Additionally, when zero is given by subtraction of severity scores, it may imply that the severity change could not be recognised by dermatologists.

The multiple linear regression model can be considered as an estimator of severity changes. The performance of the estimator is analysed through the coefficient of determination R^2 and the p-value in an F-test. The coefficient of determination R^2 analyses how the constructed linear model fits the severity changes. It indicates the percentage of variation that can be explained by the constructed linear model. It is the ratio of the regression sum of squares to the total sum of squares :

$$R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2} \tag{6.6}$$

where $\bar{y}$ is the mean value of the severity scores.

The F-test tests the significance of the estimated regression coefficients. The hypothesis is:

$$\begin{aligned} &\mathcal{H}_0 : \text{all of the regression coefficients are equal to zero} \\ &\mathcal{H}_1 : \text{at least one regression coefficient is not zero} \end{aligned} \tag{6.7}$$

If the p-value is less than 0.05, the null hypothesis is rejected and the model is significant.

The rest of this section separately describes the experimental results of the erythema severity change analysis and the scaling severity change analysis.

6.2.1 Experiment on the erythema severity changes

Table 6.1 shows the results for 17 images with erythema severity changes and Table 6.2 shows the results for 14 images with scaling severity changes. In both cases, severity change of 2, severity change of 1 and no severity changes are included.

In Table 6.1, the "score before" and "score after" are the severity scores given by dermatologists at the fist time point and the second time point respectively. Besides, the index of image pairs, the change score, and the erythema severity change features: $D(\nabla\overline{q^h})$ and $D(\nabla\overline{q^m})$ are given.

It is observed that the decrease in erythema severity is accompanied by decrease of the mean haemoglobin difference $\nabla\overline{q^h}$ and increase of the mean melanin difference $\nabla\overline{q^m}$. When the erythema severity does not change, both $\nabla\overline{q^h}$ and $\nabla\overline{q^m}$ may increase or decrease. Furthermore, there is no clear relationship between values of the severity change features and the erythema severities at the start time point.

The multiple linear regression model is given by:

$$S_{\text{Erythema}} = 0.3653D(\nabla\overline{q^h}) - 0.0468D(\nabla\overline{q^m}) + 0.0824 \tag{6.8}$$

where S_{Erythema} is the estimated erythema severity change score obtained by using PASI. The statistical R^2 value, which indicates the correlation of the estimated severity change score and the actual score, is 0.6747. The p-value is 0.0004, which is much less than the

Table 6.1: Erythema severity change scores with the severity change features and the before-after severity scores.

Index	Score before	Score after	Changes score	$D(\nabla\overline{q^h})$	$D(\nabla\overline{q^m})$
1	4	2	-2	-3.203	9.663
2	3	1	-2	-2.466	15.795
3	3	2	-1	-2.238	4.909
4	3	2	-1	-1.925	4.909
5	3	2	-1	-1.374	0.033
6	3	2	-1	-0.773	15.515
7	4	3	-1	-0.666	17.174
8	2	2	0	-0.740	8.120
9	2	2	0	-0.598	8.238
10	2	2	0	0.203	4.547
11	2	2	0	1.243	13.112
12	2	2	0	2.784	9.848
13	3	3	0	-0.592	10.224
14	3	3	0	0.063	13.077
15	3	3	0	0.236	6.545
16	3	3	0	0.299	10.854
17	3	3	0	3.165	18.355

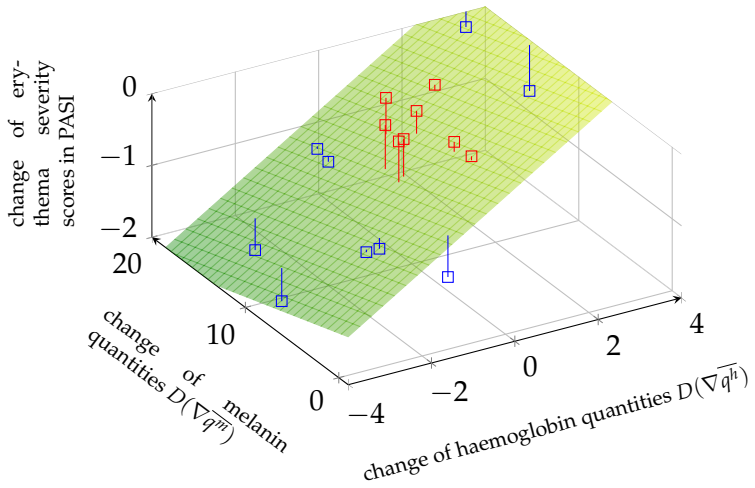


Figure 6.2: Distribution of erythema severity change features in Table 6.1 and the severity change score.

significance level of 0.05.

Figure 6.2 shows the distribution of the erythema severity change features with severity change scores. The estimated erythema severity change score S_{Erythema} is illustrated with the green-yellow plane.

The parameters of the linear regression model indicate the linear relationship between erythema severity change features and the changes in erythema severity scores. Decreasing the mean haemoglobin difference has a greater effect on the decrease of erythema severity than an increase in the mean melanin difference. This fits the fact that the haemoglobin component contributes to the redness of the skin, while the melanin component causes yellowish to the skin. The melanin component slightly effects the severity changes due to its indirect effect.

6.2.2 Experiment on the scaling severity changes

Table 6.2 shows the index of image pairs, the score before, the score after, the scaling severity change score, and the corresponding severity change features: $D(r)$ and $D(\overline{g})$. It is clear that a decrease of the scaling severity score changes with the decrease of relative scaling area r and the roughness degree $\overline{g}$, and the severity scores at the before time do not affect the change quantities of the severity features.

Table 6.2: Scaling severity change scores with the severity change features and the before-after severity scores.

Index	Score before	Score after	Change score	$D(r)$	$D(\overline{g})$
1	4	2	-2	-0.087	-0.105
2	2	1	-1	-0.512	-0.100
3	2	1	-1	-0.222	-0.237
4	2	1	-1	-0.045	-0.109
5	3	2	-1	-0.323	-0.132
6	4	3	-1	-0.164	-0.021
7	1	1	0	-0.006	0.150
8	2	2	0	0.039	-0.014
9	2	2	0	0.050	-0.218
10	3	3	0	-0.053	0.043
11	3	3	0	-0.050	0.119
12	3	3	0	0.031	-0.061
13	3	3	0	0.052	0.052
14	3	3	0	0.068	-0.133

The multiple linear regression model of scaling severity changes is given by:

$$S_{\text{Scaling}} = -0.2373 + 1.9711D(r) + 1.6613D(\overline{g}) \tag{6.9}$$

where S_{Scaling} is the estimated scaling severity change score. The R^2 of this model is 0.4255 and the p-value is 0.0474.

Figure 6.3 presents the distribution of the scaling severity change features with the severity change scores. The plane of estimated scaling severity change score S_{Scaling} is drawn with a green-yellow colour.

Observe that the coefficients of $D(r)$ and $D(\overline{g})$ in Eq. 6.9 are quite near, and this implies that the changes of scaling severities are nearly equally affected by the changes of relative scaling area and the changes of roughness degree. This observation matches the scaling scoring fact, where relative scaling area and the scaling roughness degree are considered together to decide the scaling severity (see Chapter 5).

The low R^2 value is due to the severity change features with a two-score change. The residual of this observation is -1.418, while the residuals of the rest observations

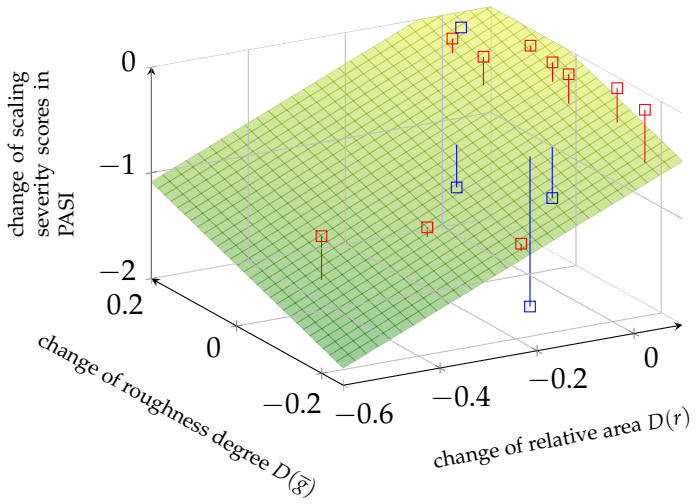


Figure 6.3: Distribution of scaling severity change features in Table 6.2 with the severity change scores.

(a) (b)

Figure 6.4: The selected pair of images with severity change of 2. (a) A psoriatic lesion with scaling severity 4; (b) The lesion in (a) evolved to the degree with scaling severity 2.

are between ± 0.5. This is because in the selected image pair the higher scaling severity score lesion with a 4 score is mostly covered by scaling that displays itself as a smooth plaque. The Gabor feature is not good at differentiating this type of roughness changes. Figure 6.4 shows the pair of the images. In Figure 6.4 (a), the lesion is covered by thick scaling and is smooth. The scaling severity of the lesion is graded 4. The same lesion in Figure 6.4 (b) is covered by fine scaling, but smooth as well. This lesion has a score 2. As observed in Figure 6.4, though the roughness of scaling does not change a lot between the two images, the colour of scaling changes dramatically. The change of scaling colour can be used in the future work to the improve the performance of the severity changes assessment method.

6.3 Summary

This chapter presents the work on the severity change analysis using the severity features proposed in Chapter 5. The multiple linear regression models of severity changes with respect to erythema and scaling are built. The reliability of the models are validated by evaluating on bitemporal psoriasis skin images and comparing the results with changes of the severity scores given by dermatologists. These models provide a potential to measure the severity changes on a continuous scale, rather than on a numerical scale as presented in PASI. The continuous scale measurement is more accurate for the assessment of small severity changes that PASI is not able to measure.

Part V

summary

Chapter 7
Conclusion and Future Work

PSORIASIS is a common chronic skin disease with no known cure. Since the treatment is individualised, it is important to compare the efficacy of various treatments in order to find the most efficacious one. Currently, commonly used evaluation methods for psoriasis treatment efficacy use only visual assessment of psoriasis severity, and are therefore subjective and can be controversial. A reliable evaluation method is in demand. The emergence of image processing techniques has lead to a number of computer-aided diagnosis systems. As we have argued in Chapter 2, the existing research is limited to plaque psoriasis and has specifically focused on segmentation.

This thesis proposes a method for reliably evaluating psoriasis more widely. In this thesis, a general psoriatic lesion is segmented through the use of features extracted from 2D images and these results are further used in the assessment of psoriasis severity. Additionally, the treatment efficacy is evaluated through the comparison of severity in a long-term treatment.

7.1 Contributions

This thesis proposes new techniques for the analysis of psoriasis treatment efficacy using only 2D images as would normally be taken during clinical treatment. The contributions of this thesis are listed as follows.

1. We have developed an erythema segmentation method based on the skin colour decomposition. The haemoglobin and melanin components are extracted through an independent component analysis, and are validated to be consistent with the

haemoglobin and melanin pigments in human skin. As presented in Chapter 3, the use of skin colour decomposition and a Support Vector Machine (SVM) for classification is shown to outperform the previous method that uses the nearest neighbour classification for segmentation [11], especially with respect to mild erythema.

2. We have proposed a novel scaling segmentation method that is able to identify the location of scaling with various morphologies and where scaling and normal skin are adjacent. In Chapter 4, two scaling features are proposed: one is the scaling contrast map enhancing the contrast between scaling and the surrounding erythema, and the other is the Gabor texture highlighting difference between the rough scaling and smooth normal skin.

   A soft-constrained k-means method is proposed to collect training sets of normal skin and scaling from the image being analysed. Doing this removes any bias from training samples and ends up with a training set calibrated to the image. Scaling pixels are identified with a Markov random filed based on the hyperplane of an SVM. The SVM initially classifies pixels as scaling or normal skin. This segmentation method includes the image structure in the segmentation process and is a non-parametric method that does not require assumptions about probability distributions in a feature space. This method is good at the segmentation of scaling under different imaging conditions and with different skin types.
3. The method for evaluation improves assessment of erythema and scaling severity. Novel severity features are proposed for erythema and scaling separately. Erythema severity is described by the relative haemoglobin and melanin quantities. Scaling severity is modelled with the relative area and roughness degree of the scaling. As demonstrated in Chapter 5, the performance of our severity features over competitors is shown. Our severity features correlate well with dermatologists' observations. Specifically, the performance is shown to be superior to the previously investigated relative colour components in the RGB colour space for erythema severity assessment and scaling area for scaling severity assessment.
4. A method for evaluation of the evolution of psoriasis in a long-term treatment

has been proposed. In Chapter 6, the severity changes are analysed through the changes in severity features and are modelled using a multiple linear regression analysis. In the multiple linear regress analysis, the degree of a severity change is indicated using a continuous scale. This study allows us to assess severity changes in a long-term treatment, and shows a potential to evaluate small severity changes that cannot be detected by Psoriasis area and severity index (PASI) alone. The method goes beyond previous studies that rely on registration of psoriatic lesions. The previous studies assume that the borders of a psoriatic lesion are kept the same in a treatment, and are therefore not applicable to a long-term treatment, since the appearance of psoriatic lesions dramatically changes over time.

Through the four contributions noted above, we addressed the three major gaps identified between existing methods and objectively diagnosing the efficacy of psoriasis treatment as pointed out in Chapter 2. The filled gaps are as follows.

- A general psoriatic lesion can be segmented through the segmentation of erythema and scaling separately. By using this strategy, segmentation of psoriasis is not limited to plaque psoriasis. The strategy is applicable to the segmentation of other types of psoriasis, such as pustular psoriasis, guttate psoriasis, flexural psoriasis and erythrodermic psoriasis.
- The accuracy of the severity assessment of psoriasis is improved by using the proposed severity features (see Chapter 5). Erythema severity is assessed through the relative haemoglobin and melanin quantities, and scaling severity is assessed through the relative roughness and relative area. The severity features are more correlated with the severity degrees than previously investigated severity features, and improve the scoring accuracy when severity classifiers are used to grade the severity.
- By using severity change features, changes of psoriasis severity can be assessed without registration of psoriasis images (see Chapter 6). This method provides a solution for the assessment of psoriasis severity changes in a long-term treatment, where borders of psoriatic lesions dramatically change and the registration of psoriasis images are not applicable.

7.2 Discussions and future work

This thesis focuses on the analysis of treatment efficacy of a target lesion. It will be beneficial in the experimental study of psoriasis treatment. When it comes to clinical practice that requires considering the percentage of psoriasis coverage of the whole body, a method for the segmentation of psoriatic lesions over the entire body is needed. It may be possible to do this by dividing the body image into blocks, and segmenting the lesions in each block. Additionally, the body area, which is used to calculate the coverage percentage, can be measured through pixel counting of the skin region. Even though there is curvature in the human body, it does not affect the evaluation of psoriasis severity [26].

Additionally, a limitation of this work is that an evaluation of the thickness intensity degree is not covered. The thickness intensity degree describing the elevation of the lesion plays an important role in the PASI index and similar indices. In the future, two kinds of methods will be investigated to evaluate the thickness intensity degree.

To do the thickness analysis, one has to analyse the texture information around the borders of a lesion in a 2D skin image. We developed a Gabor texture to analyse the roughness degree in Chapter 5. Since the roughness can be considered as a small amount of elevation, it is believed that the texture analysis research would be helpful to evaluate the thickness intensity degree. The other one is to work on 3D reconstruction techniques to reconstruct the depth information of the lesion appearance. The difficulty of reconstructing 3D images of a lesion will be in the resolution of depth, considering that the lesion borders are around 1cm high.

Several directions that can be explored in future to improve the current work are listed below:

1. The accuracy of erythema segmentation can be improved through collecting training sets from individual images and multi-window filtering. The erythema segmentation suffers disturbance from shadows and uneven skin colour. Applying this kind of techniques as what we have applied to the scaling segmentation would reduce the disturbance.
2. The scaling segmentation technique can be improved by applying hair removal algorithms to recover the skin regions covered by hair. In the scaling segmenta-

tion algorithm in Chapter 4, it is hard to differentiate normal skin from scaling when hair is present, since hair causes the surrounding normal skin to respond to Gabor filters in the same way as the scaling. A hair removal algorithm, such as the conventional Dull Razor algorithm [156], which is especially useful for removing long and thick hairs, would be helpful to eliminate the disadvantage.

3. The performance of the severity scoring method can be improved. In Chapter 5, erythema severity and scaling severity are automatically evaluated through our severity features and machine learning models. However, there is still a significant scope to improve the methods by investigating more correlated severity features and comprehensive classification models. Moreover, rather than considering the severity degrees as four categories, future work can use a multinomial regression analysis to model the severity degrees as a continuous number. Evaluating the severity with a continuous number will improve the precision of the evaluation given by human observers.
4. There is still work to do by collecting a larger set of samples to validate the severity change model. In Chapter 6, limited samples are collected to test the consistency between the proposed models and the severity changes diagnosed by the dermatologists. Even though the result is encouraging, the consistency of the model would be further validated with a much bigger dataset. Additionally, a bigger dataset will provide a better understanding between the severity change features and the severity change degrees.

7.3 Conclusion

To the best of our knowledge, this thesis presents the first work to reliably evaluate the efficacy of psoriasis treatment through the analysis of general psoriatic lesion images. It shows the potential of using a computer-aided image processing system to objectively and quantitatively evaluate the psoriasis severity. We hope that based on this work, the future research will advance the psoriasis diagnosis as well as the treatment research.

Bibliography

[1] J. F. Fowler, M. S. Duh, L. Rovba, S. Buteau, L. Pinheiro, F. Lobo, J. Sung, J. J. Doyle, A. Swensen, D. A. mallett, and G. Kosicki, "The impact of psoriasis on health care costs and patient work loss," *Journal of American Academy of Dermatology*, vol. 59, no. 5, pp. 772–780, 2008.

[2] S. Kreft, M. Kreft, A. Resman, P. Marko, and K. Z. Kreft, "Computer-aided measurement of psoriatic lesion area in a multicenter clinical trial–comparison to physician's estimations," *Journal of Dermatological Science*, vol. 44, no. 1, pp. 21–27, 2006.

[3] K. Hoffmann, T. Dirschka, H. Schwarze, S. el Gammal, U. Matthes, A. Hoffmann, and P. Altmeyer, "20 MHz sonography, colorimetry and image analysis in the evaluation of psoriasis vulgaris," *Journal of Dermatological Science*, vol. 9, no. 2, pp. 103–110, 1995.

[4] M. A. Fadzil, H. Fitriyah, E. Prakasa, H. Nugroho, S. Hussein, and A. M. Affandi, "Objective assessment of psoriasis lesion thickness for PASI scoring using 3D digital imaging," *World Academy of Science, Engineering and Technology*, vol. 63, pp. 109–115, 2010.

[5] L. Savolainen, J. Kontinen, J. Röning, and A. Oikarinen, "Application of machine vision to assess involved surface in patients with psoriasis," *British Journal of Dermatology*, vol. 137, pp. 395–400, Apr. 1997.

[6] J. Röing, R. Jacques, and J. Kontinen, "Area assessment of psoriatic lesions based on variable thresholding and subimage classification," in *Vision Interface '99*, (Trois-Riviéres, Canda), pp. 303–311, 19-21,May 1999.

[7] D. Delgado, B. Ersboll, and J. M. Carstensen, "An image based system to automatically and objectively score the degree of redness and scaling in psoriasis lesions," in *13th Danish Conference on Image Analysis and Pattern Recognition*, pp. 130–137, 2004.

[8] J. Taur, "Neuro-fuzzy approach to the segmentation of psoriasis images," *Journal of VLSI Signal Processing*, vol. 35, no. 1, pp. 19–27, 2003.

[9] J. Taur, G. Lee, C. Tao, C. Chen, and C. Yang, "Segmentation of psoriasis vulgaris images using multiresolution-based orthogonal subspace techniques," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, pp. 390–402, Apr. 2006.

[10] D. D. Gomez, C. Butakoff, B. Ersbøll, and J. M. Carstensn, "Automatic change detection and quantification of dermatological diseases with an application to psoriasis images," *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1012–1018, 2007.

[11] M. Ahmad Fadzil, H. Nugroho, S. Norashikin, and H. H. Suraiya, "Assessment of therapeutic response in skin pigment disorder treatment," in *Information Technology, 2008. ITSim 2008. International Symposium on*, vol. 1, pp. 1–8, 2008.

[12] C. Doukas, P. Stagkopoulos, C. Kiranoudis, and I. Maglogiannis, "Automated skin lesion assessment using mobile technologies and cloud platforms," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 2444–2447, Aug 2012.

[13] G. Yang, Y. Ren, Q. Pan, G. Ning, S. Gong, G. Cai, Z. Zhang, L. Li, and J. Yan, "A heart failure diagnosis model based on support vector machine," in *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, vol. 3, pp. 1105–1108, Oct 2010.

[14] G.Maletti, B. Ersboll, and K. Conradsen, "A combined alignment and registration scheme of lesions with psoriasis," *Information Sciences*, no. 175, pp. 141–159, 2005.

[15] M. Meier and P. B.Sheth, "Clinical spectrum and severity of psoriasis," *Current Problems in Dermatology*, vol. 38, pp. 1–20, 2009.

[16] Committee for medicinal products for human use, *Guideline on clinical investigation of medicinal products indicated for the treatment of psoriasis*. European Medicines Agnecy, 2004.

[17] M. A. Cimmino, "Epidemiology of psoriasis and psoriatic arthritis," *Reumatismo*, vol. 59 Suppl 1, pp. 19–24, 2007.

[18] H. S. Javitz, M. M. Ward, E. Farber, L. Nail, and S. G. Vallw, "The direct cost of care for psoriasis and psoriatic arthritis in the united states," *Journal of the American Academy of Dermatology*, vol. 46, no. 6, pp. 850–860, 2002.

[19] N. Jenner, J. Campbell, A. Plunkett, and R. Marks, "Cost of psoriasis: a study on the morbidity and financial effects of having psoriasis in Australia," *Australasian Journal of Dermatology*, vol. 43, no. 4, pp. 255–261, 2002.

[20] E. Puzenat, V. Bronsard, S. Prey, P. Gourraud, S. Aractingi, M. Bagot, B. Cribier, P. Joly, D. Jullien, M. Le Maitre, C. Paul, M. Richard-Lallemand, J. Ortonne, and F. Aubin, "What are the best outcome measures for assessing plaque psoriasis severity? a systematic review of the literature," *Journal of the European Academy of Dermatology and Venereology*, vol. 2, pp. 10–16, Apr. 2010.

[21] L. Naldi, "Scoring and monitoring the severity of psoriasis. what is the preferred method? what is the ideal method? is PASI passé? facts and controversies," *Clinics in Dermatology*, vol. 28, no. 1, pp. 67–72, 2010.

[22] B. A. Louden, D. J. Pearce, W. Lang, and S. R. Feldman, "A simplified psoriasis area severity index (SPASI) for rating psoriasis severity in clinic patients," *Dermatology Online Journal*, vol. 10, no. 2, 2004. `http://www.escholarship.org/uc/item/18w9j736`.

[23] S. Feldman and G. Krueger, "Psoriasis assessment tools in clinical trials," *Annals of the Rheumatic Diseases*, vol. 64, pp. 65–68, 2005.

[24] R. Marks, S. Barton, D. Shuttleworth, and A. Finlay, "Assessment of disease progress in psoriasis," *Archives of dermotology*, vol. 125, no. 2, pp. 235–240, 1989.

[25] A. D. Ormerod, C. M. Dwyer, R. Weller, D. H. Cox, and R. Price, "A comparison of subjective and objective measures of reduction of psoriasis with the use of ultrasound, reflectance colorimetry, computerized video image analysis, and nitric oxide production," *Journal of the American Academy of Dermatology*, vol. 37, pp. 51–57, Jul. 1997.

[26] B. Ramsay and C. M. Lawrence, "Measurement of involved surface area in patients with psoriasis," *British Journal Of Dermatology*, vol. 124, no. 6, pp. 565–570, 1991.

[27] B. Diffey, R. Oliver, and P. Farr, "A portable instrument for quantifying erythema induced by ultraviolet radiation," *British Journal of Dermatology*, vol. 111, pp. 663–672, Dec. 1984.

[28] I. H. Kwon, S. M. Woo, J. W. Choi, H. H. Kwon, and J. I. Youn, "Recovery from tanning induced by narrow-band UVB phototherapy in brown-skinned individuals with psoriasis: twelve-month follow-up," *Journal of Dermatological Science*, vol. 57, no. 1, pp. 12–18, 2010.

[29] P. Clarys, K. Alewaeters, R. Lambrecht, and A. Barel, "Skin color measurements: comparison between three instruments: the chromameter, the dermaspectrometer and the mexameter," *Skin Research and Technology*, vol. 6, no. 4, pp. 230–238, 2000.

[30] J. Zwinkels, "Colour-measuring instruments and their calibration," *Displays*, vol. 16, no. 4, pp. 163–171, 1996.

[31] J. Serup and T. Agner, "Colorimetric quantification of erythema–a comparison of two colorimeters (Lange Micro Color and Minolta Chroma Meter CR-200) with a clinical scoring scheme and laser-doppler flowmetry," *Clinical And Experimental Dermatology*, vol. 15, no. 4, pp. 267–272, 1990.

[32] M. H. Ahmad Fadzil, D. Ihtatho, A. M. Affandi, and S. Hussein, "Objective assessment of psoriasis erythema for PASI scoring," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 4070 –4073, Aug. 2008.

[33] L. Vaillant, M. Berson, L. Machet, A. Callens, L. Pourcelot, and G. Lorette, "Ultrasound imaging of psoriatic skin: a noninvasive technique to evaluate treatment of psoriasis," *International Journal Of Dermatology*, vol. 33, no. 11, pp. 786–790, 1994.

[34] H. Morsy, S. Kamp, L. Thrane, N. Behrendt, B. Saunder, H. Zayan, E. A. Elmagid, and G. B. E. Jemec, "Optical coherence tomography imaging of psoriasis vulgaris: correlation with histology and disease severity," *Archives of Dermatological Research*, vol. 302, no. 2, pp. 105–111, 2009.

[35] M. Alper, A. Kavak, A. Parlak, R. Demirci, I. Belenli, and N. Yesildal, "Measurement of epidermal thickness in a patient with psoriasis by computer-supported image analysis," *Brazilian Journal of Medical and Biological Research*, vol. 37, pp. 111–117, 2004.

[36] P. Singh, H. P. Soyer, W. S. Jason Wu, and S. Gilmore, "Tele-assessment of psoriasis area and severity index: a study of the accuracy of digital image capture," *Australasian Journal of Dermatology*, no. 52, pp. 259–263, 2011.

[37] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, pp. 198–211, 2007.

[38] K. Suzuki, "A review of computer-aided diagnosis in thoracic and colonic imaging," *Quantitative Imaging In Medicine And Surgery*, vol. 2, no. 3, pp. 163–176, 2012.

[39] Q. Abbas, M. Emre Celebi, and I. Fondón, "Computer-aided pattern classification system for dermoscopy images," *Skin Research and Technology*, vol. 18, no. 3, pp. 278–289, 2012.

[40] H. Masmoudi, S. Hewitt, N. Petrick, K. Myers, and M. Gavrielides, "Automated quantitative assessment of HER-2/neu immunohistochemical expression in breast cancer," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 6, pp. 916–925, 2009.

[41] M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel, and M. Gurcan, "Computerized classification of intraductal breast lesions using histopathological im-

ages," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 7, pp. 1977–1984, 2011.

[42] P. Tabrizi, S. Rezatofighi, and M.-J. Yazdanpanah, "Using PCA and LVQ neural network for automatic recognition of five types of white blood cells," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 5593–5596, 2010.

[43] J. E. Ball, T. W. Butler, and L. Bruce, "Towards automated segmentation and classification of masses in mammograms," in *Engineering in Medicine and Biology Society, 2004. EMBS 2004. 26th Annual International Conference of the IEEE*, vol. 1, pp. 1814–1817, 2004.

[44] X. Liu, J. Liu, and Z. Feng, "Mass classification in mammography with morphological features and multiple kernel learning," in *Bioinformatics and Biomedical Engineering, (ICBBE) 2011 5th International Conference on*, pp. 1–4, 2011.

[45] G.-C. Len, C.-M. Wang, and W.-J. Wang, "An unsupervised linear discriminant analysis approach to multispectral MRI images classification," in *Machine Learning and Cybernetics, 2007 International Conference on*, vol. 4, pp. 2018–2023, 2007.

[46] P. Korfiatis, A. Karahaliou, A. Kazantzi, C. Kalogeropoulou, and L. Costaridou, "Towards quantification of interstitial pneumonia patterns in lung multidetector CT," in *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*, pp. 1–5, 2008.

[47] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 7, pp. 1810–1820, 2009.

[48] G. G. Vimala and S. Mohideen, "Automatic detection of optic disk and exudate from retinal images using clustering algorithm," in *Intelligent Systems and Control (ISCO), 2013 7th International Conference on*, pp. 280–284, 2013.

[49] A. Basavanhally, S. Doyle, and A. Madabhushi, "Predicting classifier performance with a small training set: applications to computer-aided diagnosis and prognosis," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pp. 229–232, 2010.

[50] K. Nguyen, B. Sabata, and A. K. Jain, "Prostate cancer grading: gland segmentation and structural features," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 951–961, 2012.

[51] S. Zambanini, R. Sablatnig, H. Maier, and G. Langs, "Automatic image-based assessment of lesion development during hemangioma follow-up examinations," *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 83–94, 2010.

[52] D. Assefa, H. Keller, C. Menard, N. Laperriere, R. Ferrari, and I. Yeung, "Robust texture features for response monitoring of glioblastoma multiforme on T1-weighted and T2-FLAIR MR images: a preliminary investigation in terms of identification and segmentation," *Medical physics*, vol. 37, no. 4, pp. 1722–1736, 2010.

[53] H.-S. Wu, M. Fiel, T. Schiano, M. Ramer, D. Burstein, and J. Gil, "Segmentation of textured cell images based on frequency analysis," *Image Processing, IET*, vol. 5, no. 2, pp. 148–158, 2011.

[54] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 9, pp. 1661–1677, 2011.

[55] J. Lu, D. Wang, L. Shi, and P.-A. Heng, "Automatic liver segmentation in CT images based on support vector machine," in *Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on*, pp. 333–336, 2012.

[56] S. Nedevschi, A. Ciurte, and G. Mile, "Kidney CT image segmentation using multi-feature em algorithm, based on gabor filters," in *Intelligent Computer Communication and Processing, 2008. ICCP 2008. 4th International Conference on*, pp. 283–286, 2008.

[57] S. Morales, V. Naranjo, J. Angulo, and M. Alcaniz, "Automatic detection of optic disc based on PCA and mathematical morphology," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 4, pp. 786–796, 2013.

[58] P. Choorat, W. Chiracharit, and K. Chamnongthai, "A single tooth segmentation using structural orientations and statistical textures," in *Biomedical Engineering International Conference (BMEICON), 2011*, pp. 294–297, 2011.

[59] H. Ranjzad, A. Ebrahimi, and H. Sadigh, "Improving feature vectors for iris recognition through design and implementation of new filter bank and locally compound using of PCA and ICA," in *Applied Sciences on Biomedical and Communication Technologies, 2008. ISABEL '08. First International Symposium on*, pp. 1–5, 2008.

[60] A. Eguizabal, A. Laughney, P. Garcia-Allende, V. Krishnaswamy, W. Wells, K. Paulsen, B. Pogue, J. Lopez-Higuera, and O. Conde, "ICA-guided delineation of breast cancer pathology," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 1611–1614, 2012.

[61] W. Yang, H. Xia, B. Xia, L. M. Lui, and X. Huang, "ICA-based feature extraction and automatic classification of ad-related MRI data," in *Natural Computation (ICNC), 2010 Sixth International Conference on*, vol. 3, pp. 1261–1265, 2010.

[62] S. C. Park, B. Chapman, and B. Zheng, "A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on CT images: preliminary investigation," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 6, pp. 1519–1527, 2011.

[63] S. Kostopoulos, D. Cavouras, A. Daskalakis, P. Bougioukos, P. Georgiadis, G. Kagadis, I. Kalatzis, P. Ravazoula, and G. Nikiforidis, "Colour-texture based image analysis method for assessing the hormone receptors status in breast tissue sections," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 4985–4988, 2007.

[64] I. Isgum, B. van Ginneken, and M. Prokop, "A pattern recognition approach to automated coronary calcium scoring," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, pp. 746–749, 2004.

[65] I. Isgum, B. Van Ginneken, A. Rutten, and M. Prokop, "Automated coronary calcification detection and scoring," in *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pp. 1127–132, 2005.

[66] T. H. Lee, M. Fauzi, R. Komiya, and S.-C. Haw, "Unsupervised abnormalities extraction and brain segmentation," in *Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on*, vol. 1, pp. 1185–1190, 2008.

[67] C.-S. Yang and M.-Y. Lee, "Parametric data mining and diagnostic rules for digital thermographs in breast cancer," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 98–101, 2008.

[68] M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. Rudnicka, C. Owen, and S. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 9, pp. 2538–2548, 2012.

[69] L. Vibha, G. HarshaVardhan, K. Pranaw, P. Deepa Shenoy, K. Venugopal, and L. Patnaik, "Classification of mammograms using decision trees," in *Database Engineering and Applications Symposium, 2006. IDEAS '06. 10th International*, pp. 263–266, 2006.

[70] A. L. Huynen, R. J. B. Giesen, R. Laduc, F. M. J. Debruyne, and H. Wijkstra, "Hierarchical decision tree for the classification of prostate tissue," in *Engineering in Medicine and Biology Society, 1992. 14th Annual International Conference of the IEEE*, vol. 5, pp. 2100–2101, 1992.

[71] A. Pitiot, A. Toga, and P. Thompson, "Adaptive elastic segmentation of brain MRI via shape-model-guided evolutionary programming," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 8, pp. 910–923, 2002.

[72] M. Fraz, P. Remagnino, A. Hoppe, and S. Barman, "Retinal image analysis aimed at extraction of vascular structure using linear discriminant classifier," in *Computer Medical Applications (ICCMA), 2013 International Conference on*, pp. 1–6, 2013.

[73] T. Wirtti and E. O. T. Salles, "Segmentation of masses in digital mammograms," in *Biosignals and Biorobotics Conference (BRC), 2011 ISSNIP*, pp. 1–7, 2011.

[74] P. Spyridonos, D. Glotsos, D. Cavouras, P. Ravazoula, V. Zolota, and G. Nikiforidis, "Pattern recognition based segmentation method of cell nuclei in tissue section analysis," in *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*, vol. 2, pp. 1121–1124, 2002.

[75] M. Garcia, R. Hornero, C. Sanchez, M. Lopez, and A. Diez, "Feature extraction and selection for the automatic detection of hard exudates in retinal images," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 4969–4972, 2007.

[76] A. Marcano-Cedeño, A. Alvarez-Vellisco, and D. Andina, "Artificial metaplasticity MLP applied to image classification," in *Industrial Informatics, 2009. INDIN 2009. 7th IEEE International Conference on*, pp. 650–653, 2009.

[77] N. Mohamad, F. Zaini, A. Johari, I. Yassin, and A. Zabidi, "Comparison between levenberg-marquardt and scaled conjugate gradient training algorithms for breast cancer diagnosis using MLP," in *Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on*, pp. 1–7, 2010.

[78] W. Xu, W. Liu, L. Li, G. Shao, and J. Zhang, "Identification of masses and microcalcifications in the mammograms based on three neural networks: comparison and discussion," in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pp. 2299–2302, 2008.

[79] B. Lerner, M. Levinstein, B. Rosenberg, H. Guterman, I. Dinstein, and Y. Romem, "Feature selection and chromosome classification using a multilayer perceptron neural network," in *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*, vol. 6, pp. 3540–3545, 1994.

[80] D. W. K. Wong, J. Liu, J. H. Lim, N.-M. Tan, Z. Zhang, H. Li, S. Lu, and T. Y. Wong, "Method of detecting kink-bearing vessels in a retinal fundus image," in *Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on*, pp. 1690–1694, 2010.

[81] K. Goatman, A. Fleming, S. Philip, G. Williams, J. Olson, and P. Sharp, "Detection of new vessels on the optic disc using retinal photographs," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 4, pp. 972–979, 2011.

[82] N. Zhang, S. Ruan, S. Lebonvallet, Q. Liao, and Y. Zhu, "Multi-kernel svm based classification for brain tumor segmentation of MRI multi-sequence," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 3373–3376, 2009.

[83] A. Sathya, S. Senthil, and A. Samuel, "Segmentation of breast MRI using effective fuzzy C-Means method based on support vector machine," in *Information and Communication Technologies (WICT), 2012 World Congress on*, pp. 67–72, 2012.

[84] P. Li, K.-L. Chan, S.-M. Krishnan, and Y. Gao, "Detecting abnormal regions in colonoscopic images by patch-based classifier ensemble," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, pp. 774–777, 2004.

[85] Y. Fu, W. Zhang, M. Mandal, and M. Meng, "Computer-aided bleeding detection in WCE video," *Biomedical and Health Informatics, IEEE Journal of*, vol. PP, no. 99, pp. 1–7, 2013.

[86] M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, D. Cavouras, and S. Theodoridis, "Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers," *Artificial intelligence in medicine*, vol. 37, pp. 145–162, 06 2006.

[87] K. Chan, T.-W. Lee, P. Sample, M. Goldbaum, R. Weinreb, and T. Sejnowski, "Comparison of machine learning and traditional classifiers in glaucoma diagnosis," *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 9, pp. 963–974, 2002.

[88] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonca, "Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests," *BMC Research Notes*, vol. 4, no. 1, pp. 299–312, 2011.

[89] E. Frias-Martinez, A. Sanchez, and J. Velez, "Support vector machines versus multilayer perceptrons for efficient off-line signature recognition," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 6, pp. 693–704, 2006.

[90] C. Schwarz, A. Tsui, E. Fletcher, B. Singh, C. DeCarli, and O. Carmichael, "Impact of Markov random field optimizer on MRI-based tissue segmentation in the aging brain," in *Engineering in Medicine and Biology Society (EMBC), 2011 Annual International Conference of the IEEE*, pp. 7812–7815, 2011.

[91] L. Lihua, L. Jiangli, L. Deyu, and W. Tianfu, "Segmentation of medical ultrasound image based on Markov random field," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, pp. 968–971, 2007.

[92] Z. Ma, H. Jiang, B. Yang, and L. Zhang, "Unsupervised abdomen CT image segmentation using variable weight mrf in spatial and wavelet domain," in *Computer Sciences and Convergence Information Technology (ICCIT), 2011 6th International Conference on*, pp. 915–921, 2011.

[93] D. Ihtatho, M. Fadzil, A. Affandi, and S. Hussein, "Area assessment of psoriasis lesion for PASI scoring," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 3446 –3449, Aug. 2007.

[94] D. Ihtatho, M. Ahmad Fadzil, A. Mohd Affandi, and S. Hussein, "Automatic PASI area scoring," in *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on*, pp. 819 –822, Nov. 2007.

[95] A. Caliman and M. Ivanovici, "Psoriasis image analysis using color lacunarity," in *Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on*, pp. 1401–1406, 2012.

[96] A. Caliman, M. Ivanovici, and N. Richard, "Fractal feature-based color image segmentation for a healthcare application in dermatology," in *E-Health and Bioengineering Conference (EHB), 2011*, pp. 1–4, 2011.

[97] F. Bogo, M. Samory, A. Fortina, S. Piaserico, and E. Peserico, "Psoriasis segmentation through chromatic regions and geometric active contours," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 5388–5391, 2012.

[98] M. Nischik and C. Forster, "Analysis of skin erythema using true-color images," *Medical Imaging, IEEE Transactions on*, vol. 16, no. 6, pp. 711–716, 1997.

[99] E. Roullot, J.-E. Autegarden, P. Devriendt, and F. Leynadier, "Segmentation of erythema from skin photographs for assisted diagnosis in allergology," in *Pattern Recognition and Image Analysis*, vol. 3687 of *Lecture Notes in Computer Science*, pp. 754–763, Springer Berlin Heidelberg, 2005.

[100] K. Shang, L. Ying, N. Hai-jing, and L. Yu-fu, "Method of reducing dimensions of segmentation feature parameter applied to skin erythema image segmentation," in *Engineering in Medicine and Biology Society, 2005. EMBS 2005. 27th Annual International Conference of the*, pp. 3422 –3424, Jul. 2005.

[101] H. Iyatomi, H. Oka, M. Hagiwara, A. Miyake, M. Kimoto, K. Ogawa, and M. Tanaka, "Computerized quantification of psoriasis lesions with colour calibration: preliminary results," *Clinical and Experimental Dermatology*, vol. 34, no. 7, pp. 830–833, 2009.

[102] D. D. Gómez, B. K. Ersbøll, and J. M. Carstensen, "S.H.A.R.P: a smart hierarchical algorithm to register psoriasis," in *International workshop on Systems, Signals and Image Processing*, pp. 43–46, Sept. 2004.

[103] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, pp. 1106–1122, Mar. 2007.

[104] D. Chai and K. Ngan, "Face segmentation using skin-color map in videophone applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 4, pp. 551–564, 1999.

[105] S. Phung, A. Bouzerdoum, and S. Chai, D., "Skin segmentation using color pixel classification: analysis and comparison," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 1, pp. 148–154, 2005.

[106] W.-H. Lai and C.-T. Li, "Skin colour-based face detection in colour images," in *Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on*, pp. 56–61, 2006.

[107] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, pp. 81–96, Jan. 2002.

[108] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

[109] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain, "Face detection in color images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 696–706, 2002.

[110] A. Cheddad, J. Condell, K. Curran, and P. McKevitt, "Skin tone based steganography in video files exploiting the ycbcr colour space," in *Multimedia and Expo, 2008 IEEE International Conference on*, pp. 905–908, 2008.

[111] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in ycbcr color space," in *TENCON 2000. Proceedings*, vol. 2, pp. 421–424, 2000.

[112] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. I. Addison-Wesley, 1992.

[113] N. Tsumura, H. Haneishi, and Y. Miyake, "Independent component analysis of skin color image," *Journal of the Optical Society of America A*, vol. 16, no. 9, pp. 2169–2176, 1999.

[114] D. W. Ball, *Field Guide to Spectroscopy*. SPIE Press, 2006.

[115] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626–634, 1999.

[116] V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 709–715, 1963.

[117] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceesings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152, 1992.

[118] S. Xu, X. Ye, Y. Wu, F. Giron, J.-L. Leveque, and B. Querleux, "Automatic skin decomposition based on single image," *Computer Vision and Image Understanding*, vol. 110, no. 1, pp. 1–6, 2008.

[119] A. Krishnaswamy and G. V. G. Baranoski, "A study on skin optics," tech. rep., Natural Phenomena Simulation Gropu, School of Computer Science, University of Waterloo, Canada, Jan. 2004.

[120] C. Chang and C. Lin, *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[121] R. Hunt, *The Reproduction of Colour*, vol. I. John Wiley & Sons Ltd, 2004.

[122] M. Tkalcic and J. Tasic, "Colour spaces: perceptual, historical and applicational background," in *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, vol. 1, pp. 304–308 vol.1, 2003.

[123] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," *Lecture Notes in Computer Science*, no. 5008, pp. 66–78, 2008.

[124] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–511–I–518, 2001.

[125] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition*. The MIT Press, 2 ed., 2001.

[126] S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *Image Processing, IEEE Transactions on*, vol. 11, pp. 1160–1167, 2002.

[127] L. Ma and R. C. Staunton, "Optimum gabor filter design and local binary patterns for texture segmentation," *Pattern Recognition Letters*, vol. 29, pp. 664–672, 2008.

[128] B. Zhu, L. Jiang, Y. Luo, and Y. Tao, "Gabor feature-based apple quality inspection using kernel principal component analysis," *Journal of Food Engineering*, vol. 81, no. 4, pp. 741–749, 2007.

[129] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 572–581, 2004.

[130] D. Gabor, "Theory of communication. part 1: The analysis of information," *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, vol. 93, no. 26, pp. 429–441, 1946.

[131] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, no. 10, pp. 847–856, 1980.

[132] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[133] P. Kruizinga and N. Petkov, "Nonlinear operator for oriented texture," *Image Processing, IEEE Transactions on*, vol. 8, pp. 1395–1407, Oct. 1999.

[134] P. Kruizinga, N. Petkov, and S. Grigorescu, "Comparison of texture features based on Gabor filters," in *International Conference on Image Analysis and Processing*, pp. 142–147, 1999.

[135] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, vol. 22, no. 5, pp. 545–559, 1982.

[136] N. Petkov and P. Kruizinga, "Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells," *Biological Cybernetics*, vol. 76, no. 2, pp. 83–96, 1997.

[137] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

[138] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recognition*, vol. 24, pp. 1167–1186, 1991.

[139] K. Wagstaff, *Intelligent clustering with instance-level constraints*. PhD thesis, Cornell University, 2002.

[140] Z. Kato and T. chuen Pong, "A Markov random field image segmentation model for color textured images," *Image and Vision Computing*, vol. 24, pp. 1103–1114, 2006.

[141] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 6, pp. 1068–1080, 2008.

[142] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[143] S. Li, *Markov Random Field Models in Computer Vision*. Springer-Verlag, 1995.

[144] T. Sergios and K. Koutroumbas, *Pattern Recognition*. Academic Press, 4th edition ed., 2009.

[145] D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 51, no. 2, pp. 271–279, 1989.

[146] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004.

[147] L. Zhang, "Hierarchical block-based disparity estimation using mean absolute difference and dynamic programming," in *VLBV01: International Workshop on Very Low Bitrate Video Coding*, pp. 114–118, 2001.

[148] T. Malisiewicz and A. A. Efros, "Improving spatial support for objects via multiple segmentations," in *British Machine Vision Conference (BMVC)*, Sept. 2007.

[149] M.-C. Su and C.-H. Chou, "A modified version of the k-means algorithm with a distance based on cluster symmetry," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, pp. 674–680, Jun. 2001.

[150] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *Medical Imaging, IEEE Transactions on*, vol. 21, pp. 193–199, Mar. 2002.

[151] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

[152] N. Pal and J. Bezdek, "On cluster validity for the fuzzy c-means model," *Fuzzy Systems, IEEE Transactions on*, vol. 3, no. 3, pp. 370–379, 1995.

[153] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[154] J. Gareth, W. Daniela, H. Trevor, and T. Robert, *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.

[155] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[156] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor: a software approach to hair removal from images," *Computers in Biology and Medicine*, vol. 27, pp. 533–543, Nov. 1997.